

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

UTILITY  
PATENT APPLICATION  
TRANSMITTAL

(Only for new nonprovisional applications under 37 CFR 1.53(h))

Attorney Docket No.	06271.0024.DVUS01
First Named Inventor or Application Identifier	Gregory P. WINTER
Title	A New Method for Tapping the Immunological Repertoire
Express Mail Label No.	EL615207731US

## APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents

ADDRESS TO: Assistant Commissioner for Patents  
Box Patent Application  
Washington, DC 20231

1. ☒ \*Fee Transmittal Form (Form PTO-1082)  
(Submit an original and a duplicate for fee processing)
2. ☒ Specification [Total Pages 98]  
(preferred arrangement set forth below)
- Descriptive title of the Invention
  - Cross References to Related Applications
  - Statement Regarding Fed sponsored R&D
  - Reference to Microfiche Appendix
  - Background of the Invention
  - Brief Summary of the Invention
  - Brief Description of the Drawings (if filed)
  - Detailed Description
  - Claims
  - Abstract of the Disclosure
3. ☒ Drawing(s) (35 USC 113) [Total Sheets 19]
4. Oath or Declaration [Total Pages ]
- a. ☐ Newly executed (original or copy)
- b. ☐ Copy from a prior application (37 CFR 1.63(d))  
(for continuation/divisional with Box 17 completed)  
[Note Box 5 below]
1. ☐ **DELETION OF INVENTOR(S)**  
Signed statement attached deleting inventor(s) named  
in the prior application, see 37 CFR 1.63(d)(2) and  
1.33(b).
- ☒ Incorporation By Reference (useable if Box 4b is checked)  
The entire disclosure of the prior application is considered as  
being part of the disclosure of the accompanying application and  
is hereby incorporated by reference therein.
6. ☐ Microfiche Computer Program (Appendix)
7. Nucleotide and/or Amino Acid Sequence Submission  
(if applicable, all necessary)
- a. ☐ Computer Readable Copy
- b. ☐ Paper Copy (identical to computer copy)
- c. ☐ Statement verifying identity of above  
copies
- ACCOMPANYING APPLICATION PARTS**
8. ☐ Assignment Papers (cover sheet & document(s))
9. ☐ 37 CFR 3.73(b) Statement ☐ Power of Attorney  
(when there is an assignee)
10. ☐ English Translation Document (if applicable)
11. ☐ Information Disclosure Statement (IDS)/PTO-1449 ☐ Copies of IDS Citations
12. ☒ Preliminary Amendment
13. ☒ Return Receipt Postcard (MPEP 503)  
(should be specifically itemized)
14. ☐ \*Small Entity Statement(s) ☐ Statement filed in prior application, Status still proper and desired
15. ☐ Certified Copy of Priority Document(s)  
(if foreign priority is claimed)
16. ☐ Other:
- \*NOTE FOR ITEMS 1 & 14 IN ORDER TO BE ENTITLED TO PAY SMALL ENTITY FEES, A SMALL ENTITY STATEMENT IS REQUIRED (37 CFR § 1.27) EXCEPT IF ONE FILED IN A PRIOR APPLICATION IS RELIED UPON (37 CFR § 1.28)

## 17. If a CONTINUING APPLICATION, check appropriate box and supply the requisite information

☐ Continuation ☒ Divisional ☐ Continuation-in-part (CIP) of prior application No. 07/933,958/August 21, 1992

Prior Application Information: Examiner: J. Ketter Group/Art Unit: 1636

## 18. CORRESPONDENCE ADDRESS

☒ Customer Number or Bar Code Label 22930 or ☐ Correspondence address below  
(Insert Customer No. or Attach bar code label here)

NAME	Albert P. Halluin				
	HOWREY SIMON ARNOLD & WHITE, LLP				
ADDRESS	Box No. 34				
	1299 Pennsylvania Avenue, N.W.				
CITY	Washington	STATE	DC	ZIP CODE	20004-2402
COUNTRY	US	TELEPHONE	202-783-0800	FAX	202-383-7195
Name (Print/Type)	Albert P. Halluin		Registration No. (Attorney/Agent)	27,227	
Signature	<i>Albert P. Halluin</i>		Date	November 28, 2000	

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Box Patent Application, Washington, DC 20231

a check # 355

FORM PTO-1082

HOWREY SIMON ARNOLD &amp; WHITE, LLP

Box No. 34

1299 Pennsylvania Avenue, N.W.

Washington, D.C. 20004-2402

(650) 463-8100

Attorney Docket No. 06271.0024.DVUS01**Box Patent Application**

COMMISSIONER FOR PATENTS

Washington, D.C. 20231

Sir:

Transmitted herewith for filing is the patent application of Gregory P. WINTER; Lutz RIECHMAN; William D. HUSE; Joseph A. SORGE; and Richard A. LERNER for A NEW METHOD FOR TAPPING THE IMMUNOLOGICAL REPERTOIRE.

Also, enclosed are:

1. Preliminary Amendment (4 pages)
2. 19 sheets of drawings
3. Return postcard.

The filing fee has been calculated as shown below:

	(Col. 1)	(Col. 2)
FOR	NO. FILED	NO. EXTRA
BASIC FEE		
TOTAL CLAIMS	6-20 =	0
INDEP. CLAIMS	3-3 =	0
<input type="checkbox"/> MULTIPLE DEPENDENT CLAIM PRESENTED		

\*If the difference in Col. 1 is less than zero, enter "0" in Col. 2

SMALL ENTITY	
RATE	FEE
	\$ 355.00
x 9 =	0.00
x 40 =	0.00
+ 135 =	0.00
TOTAL	\$0.00

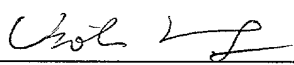
OTHER THAN A SMALL ENTITY	
RATE	FEE
	\$ 710.00
X 18 =	
x 80 =	
+ 270 =	
TOTAL	710.00

Applicant(s) hereby authorize(s) the U.S. Patent and Trademark Office (PTO) to charge payment of any fee associated with this filing to our Deposit Account No. 08-3038 referencing docket number 06271.0024.DVUS01.

Also, this application is a DIVISIONAL of U.S. Patent Application Serial No. 07/933,958, filed August 21, 1992; which is a CONTINUATION of U.S. Patent Application Serial No. 07/799,770, filed November 27, 1991; which is a CONTINUATION of U.S. Patent Application Serial No. 07/533,103, filed June 4, 1990; which is a CONTINUATION-IN-PART of U.S. Patent Application Serial No. 07/353,235, filed May 16, 1989 (Abandoned); and is a CONTINUATION-IN-PART of U.S. Patent Application Serial No. 07/353,241, filed May 17, 1989 (Abandoned).

The U.S. Patent and Trademark Office is hereby authorized to charge any fee deficiency, or credit any overpayment, to our Deposit Account No. 08-3038. A duplicate copy of this sheet is attached.

Date November 28, 2000

  
 Albert P. Halluin (25,227)  
 Viola T. Kung (Reg. No. 41,131)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Gregory P. Winter, *et al.*

Application Serial No. To be Assigned

Filed: Herewith

For: **A NEW METHOD FOR TAPPING THE  
IMMUNOLOGICAL REPERTOIRE**

Art Unit: To be Assigned

Examiner: To be Assigned

Attorney's Docket No:  
06271.0024.DVUS01

**PRELIMINARY AMENDMENT**

**Box Patent Application**

Commissioner for Patents

Washington, D.C. 20231

Sir:

Applicant respectfully requests entry of the following amendments and consideration of the following remarks prior to examination of the above-referenced patent application.

**AMENDMENTS**

**In the Specification**

Page 1, line 7, after: "This," insert -- application is a divisional of U.S. Application No. 07/933,958, filed August 21, 1992; which is a continuation of U.S. Application No. 07/799,770, filed November 27, 1991; which is a continuation of U.S. Application No. 07/533,103, filed June 4, 1990; which --.

Page 1, line 9, after "1989," insert --abandoned,--.

Page 1, line 11, after "1989," insert --abandoned,--.

**In the Claims**

Cancel Claims 1-15 and add the following new claims.

-- 16. A  $V_H$  or  $V_L$  polypeptide isolated by the method comprising the steps of:

(a) synthesizing a  $V_H$  or  $V_L$ -coding gene library containing a plurality of different  $V_H$  or  $V_L$ -coding DNA sequences by a method comprising the steps of:

(i) preparing a polynucleotide containing composition, wherein at least a portion of the polynucleotide in said composition comprise a plurality of different  $V_H$  or  $V_L$ -coding sequences;

(ii) amplifying a plurality of  $V_H$  or  $V_L$ -coding sequences in said polynucleotide containing composition;

(iii) joining in operable combination  $V_H$  or  $V_L$ -coding sequences from said  $V_H$  or  $V_L$ -coding gene library into expression vectors so as to be able to express a  $V_H$  or  $V_L$ -coding sequence, whereby a diverse library is formed;

(b) selecting and isolating from said diverse library at least one expression vector capable of producing  $V_H$  or  $V_L$  polypeptides having the desired specificity;

(c) transforming a host cell with said expression vector; and

(d) isolating a  $V_H$  or  $V_L$  polypeptide encoded by said vector from said host cell.

17. A  $V_H$  or  $V_L$  polypeptide according to Claim 16, wherein said  $V_H$  or  $V_L$  polypeptide is capable of having a catalytic activity.

18. A  $V_H$  or  $V_L$  polypeptide isolated by a method comprising the steps of:

(a) preparing a polynucleotide containing composition, wherein at least a portion of the polynucleotide in said composition comprise a plurality of  $V_H$  or  $V_L$ -coding sequences;

(b) amplifying a plurality of  $V_H$  or  $V_L$ -coding sequences from said polynucleotide containing composition by a method of amplification comprising the step of adding primer sequences capable of hybridizing upstream and downstream from a plurality of said  $V_H$  or  $V_L$ -coding sequences under conditions permitting hybridization to occur, whereby a plurality of amplified  $V_H$  or  $V_L$ -coding sequences are produced;

(c) joining in operable combination,  $V_H$  or  $V_L$ -coding sequences from said  $V_H$  or  $V_L$ -coding gene library with an expression vector so as to be able to express a  $V_H$  or  $V_L$ -coding sequence from each vector, whereby a diverse library is formed;

(d) selecting from said diverse library at least one expression vector capable of

producing a V<sub>H</sub> or V<sub>L</sub> polypeptide having a desired specificity;

- (e) transforming a host cell with said expression vector; and
- (f) isolating a V<sub>H</sub> or V<sub>L</sub> polypeptide encoded by said vector from said host cell.

19. A V<sub>H</sub> or V<sub>L</sub> polypeptide according to Claim 18, wherein said V<sub>H</sub> or V<sub>L</sub> polypeptide is capable of having a catalytic activity.

20. A V<sub>H</sub> or V<sub>L</sub> polypeptide isolated by the method comprising the steps of:

(a) producing V<sub>H</sub> or V<sub>L</sub>-coding genetic library, by a method comprising the steps of:

(i) adding a first primer, wherein said first primer is capable of hybridizing to a first conserved nucleotide sequence substantially adjacent to a plurality of V<sub>H</sub> or V<sub>L</sub> coding regions, and said coding sequences are present in a polynucleotide containing composition that comprises a plurality of different V<sub>H</sub> or V<sub>L</sub>-coding sequences;

(ii) adding a second primer to said nucleotide containing composition, wherein said second primer is capable of hybridizing to a second conserved nucleotide sequence substantially adjacent to a plurality of V<sub>H</sub> or V<sub>L</sub> coding regions and said second conserved nucleotide sequence;

(b) joining in operable combination, V<sub>H</sub> or V<sub>L</sub>-coding sequences from said V<sub>H</sub> or V<sub>L</sub>-coding gene library with expression vectors so as to be able to express a V<sub>H</sub> or V<sub>L</sub>-coding sequence from each vector, whereby a diverse library is formed;

(c) selecting and isolating from said diverse library at least one expression vector capable of producing V<sub>H</sub> or V<sub>L</sub> polypeptides having the desired specificity;

- (d) transforming a host cell with said expression vector; and
- (e) isolating a V<sub>H</sub> or V<sub>L</sub> polypeptide encoded by said vector from said host cell.

21. A V<sub>H</sub> or V<sub>L</sub> polypeptide according to Claim 20, wherein said V<sub>H</sub> or V<sub>L</sub> polypeptide is capable of having a catalytic activity. --.

### REMARKS

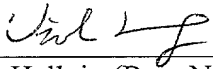
The above amendments to the claims are made for the purpose of more clearly defining what Applicants regard as the invention. New Claims 16-21 correspond to Claims 26, 27, 40, 41, 52 and 53 submitted in the Preliminary Amendment dated March 31, 1993 in

the patent application 07/933,958; those claims were canceled in response to Restriction Requirement dated September 24, 1993.

No new matter is added to any of the above Amendments. The Examiner is requested to enter the amendments prior to examination of the application.

Respectfully submitted,

Date: November 28, 2000

  
\_\_\_\_\_  
Albert P. Halluin (Reg. No. 25,227)  
Viola T. Kung (Reg. No. 41,131)

**HOWREY SIMON ARNOLD & WHITE, LLP**

Box No. 34

1299 Pennsylvania Avenue, N.W.

Washington, D.C. 20004-2402

(650) 463-8100

A NEW METHOD FOR TAPPING  
THE IMMUNOLOGICAL REPERTOIRE

Description

Cross Reference to Related Application

This is a continuation-in-part application of copending application Serial Number 353,235 having the same title and filed May 16, 1989, and Serial Number 353,241, having the same title and filed May 17, 1989, the disclosures of which are hereby incorporated by reference.

Technical Field

The present invention relates to a method for isolating a gene coding for a receptor having a preselected activity.

Background

The immune system of a mammal is one of the most versatile biological systems as probably greater than  $1.0 \times 10^7$  antibody specificities can be produced. Indeed, much of contemporary biological and medical research is directed toward tapping this repertoire. During the last decade there has been a dramatic increase in the ability to harness the output of the vast immunological repertoire. The development of the hybridoma methodology by Kohler and Milstein has made it possible to produce monoclonal antibodies, i.e., a composition of antibody molecules of a single specificity, from the repertoire of antibodies induced during an immune response.

Unfortunately, current methods for generating monoclonal antibodies are not capable of

efficiently surveying the entire antibody response induced by a particular immunogen. In an individual animal there are at least 5-10,000 different B-cell clones capable of generating unique antibodies to a small relatively rigid immunogens, such as, for example dinitrophenol. Further, because of the process of somatic mutation during the generation of antibody diversity, essentially an unlimited number of unique antibody molecules may be generated. In contrast to this vast potential for different antibodies, current hybridoma methodologies typically yield only a few hundred different monoclonal antibodies per fusion.

Other difficulties in producing monoclonal antibodies with the hybridoma methodology include genetic instability and low production capacity of hybridoma cultures. One means by which the art has attempted to overcome these latter two problems has been to clone the immunoglobulin-producing genes from a particular hybridoma of interest into a procaryotic expression system. See, for example, Robinson et al., PCT Publication No. WO 89/0099; Winter et al., European Patent Publication No. 0239400; Reading, U.S. Patent No. 4,714,681; and Cabilly et al., European Patent Publication No. 0125023.

The immunologic repertoire of vertebrates has recently been found to contain genes coding for immunoglobulins having catalytic activity. Tramontano et al., Sci., 234:1566-1570 (1986); Pollack et al., Sci., 234:1570-1573 (1986); Janda et al., Sci., 241:1188-1191 (1988); and Janda et al., Sci., 244:437-440 (1989). The presence of, or the ability to induce the repertoire to produce, antibodies molecules capable of a catalyzing chemical reaction, i.e., acting like enzymes, had previously



been postulated almost 20 years ago by W. P. Jencks in Catalysis in Chemistry and Enzymology, McGraw-Hill, N.Y. (1969).

5 It is believed that one reason the art failed to isolate catalytic antibodies from the immunological repertoire earlier, and its failure to isolate many to date even after their actual discovery, is the inability to screen a large portion of the repertoire for the desired activity. Another reason is believed to be the bias of currently available screening techniques, such as the hybridoma technique, towards the production high affinity antibodies inherently designed for participation in the process of neutralization, as opposed to catalysis.

#### Brief Summary of the Invention

20 The present invention provides a novel method for screening a larger portion of the immunological repertoire for receptors having a preselected activity than has heretofore been possible, thereby overcoming the before-mentioned inadequacies of the hybridoma technique.

25 In one embodiment, an immunoglobulin heavy chain variable region  $V_H$  gene library containing a substantial portion of the  $V_H$  gene repertoire of a vertebrate is synthesized. In preferred embodiments, the  $V_H$ -coding gene library contains at least about  $10^3$ , preferably at least about  $10^4$  and more preferably at least about  $10^5$  different  $V_H$ -coding nucleic acid strands referred to herein as  $V_H$ -coding DNA homologs.

35 The gene library can be synthesized by either of two methods, depending on the starting material.

Where the starting material is a plurality of  $V_H$ -coding genes, the repertoire is subjected to two distinct primer extension reactions. The first primer extension reaction uses a first polynucleotide synthesis primer capable of initiating the first reaction by hybridizing to a nucleotide sequence conserved (shared by a plurality of genes) within the repertoire. The first primer extension reaction produces a plurality of different  $V_H$ -coding homolog complements (nucleic acid strands complementary to the genes in the repertoire).

The second primer extension reaction produces, using the complements as templates, a plurality of different  $V_H$ -coding DNA homologs. The second primer extension reaction uses a second polynucleotide synthesis primer that is capable of initiating the second reaction by hybridizing to a nucleotide sequence conserved among a plurality of  $V_H$ -coding gene complements.

Where the starting material is a plurality of complements of different  $V_H$ -coding genes provided by a method other than the first primer extension reaction, the repertoire is subjected to the above-discussed second primer extension reaction. That is, where the starting material is a plurality of different  $V_H$ -coding gene complements produced by a method such as denaturation of double strand genomic DNA, chemical synthesis and the like, the complements are subjected to a primer extension reaction using a polynucleotide synthesis primer that hybridizes to a plurality of the different  $V_H$ -coding gene complements provided. Of course, if both a repertoire of  $V_H$ -coding genes and their complements are present in the starting material, both approaches can be used in combination.

A  $V_H$ -coding DNA homolog, i.e., a gene coding for a receptor capable of binding the preselected ligand, is then segregated from the library to produce the isolated gene. This is typically accomplished by operatively linking for expression a plurality of the different  $V_H$ -coding DNA homologs of the library to an expression vector. The  $V_H$ -expression vectors so produced are introduced into a population of compatible host cells, i.e., cells capable of expressing a gene operatively linked for expression to the vector. The transformants are cultured under conditions for expressing the receptor coded for by the  $V_H$ -coding DNA homolog. The transformants are cloned and the clones are screened for expression of a receptor that binds the preselected ligand. Any of the suitable methods well known in the art for detecting the binding of a ligand to a receptor can be used. A transformant expressing the desired activity is then segregated from the population to produce the isolated gene.

In another embodiment, the present invention contemplates a gene library comprising an isolated admixture of at least about  $10^3$ , preferably at least about  $10^4$  and more preferably at least  $10^5$   $V_H$ - and/or  $V_L$ -coding DNA homologs, a plurality of which share a conserved antigenic determinant. Preferably, the homologs are present in a medium suitable for in vitro manipulation, such as water, phosphate buffered saline and the like, which maintains the biological activity of the homologs.

A receptor having a preselected activity, preferably catalytic activity, produced by a method of the present invention, preferably a  $V_H$  or  $F_V$  as described herein, is also contemplated.

### Brief Description of the Drawings

In the drawings forming a portion of this disclosure:

Figure 1 Illustrates a schematic diagram of the immunoglobulin molecule showing the principal structural features. The circled area on the heavy chain represents the variable region ( $V_H$ ), a polypeptide containing a biologically active (ligand binding) portion of that region, and a gene coding for that polypeptide, are produced by the methods of the present invention. Sequences L03, L35, L47 and L48 could not be classified into any predefined subgroups.

Figure 2A Diagrammatic sketch of an H chain of human IgG (IgG1 subclass). Numbering is from the N-terminus on the left to the C-terminus on the right. Note the presence of four domains, each containing an intrachain disulfide bond (S-S) spanning approximately 60 amino acid residues. The symbol CHO stands for carbohydrate. The V region of the heavy (H) chain ( $V_H$ ) resembles  $V_L$  in having three hypervariable CDR (not shown).

Figure 2B Diagrammatic sketch of a human K chain (Panel 1). Numbering is from the N-terminus on the left to the C-terminus on the right. Note the intrachain disulfide bond (S-S) spanning about the same number of amino acid residues in the  $V_L$  and  $C_L$  domains. Panel 2 shows the locations of the complementarity-determining regions (CDR) in the  $V_L$  domain. Segments outside the CDR are the framework segments (FR).

Figure 3 Amino acid sequence of the  $V_H$  regions of 19 mouse monoclonal antibodies with specificity for phosphorylcholine. The designation HP indicates that the protein is the product of a

hybridoma. The remainder are myeloma proteins.

(From Gearhart et al., Nature, 291:29, 1981.)

Figure 4 Illustrates the results obtained from PCR amplification of mRNA obtained from the spleen of a mouse immunized with FITC. Lanes R17-R24 correspond to amplification reactions with the unique 5' primers (2-9, Table 1) and the 3' primer (12, Table 1), R16 represents the PCR reaction with the 5' primer containing inosine (10, Table 1) and 3' primer (12, Table 1). Z and R9 are the amplification controls; control Z involves the amplification of  $V_H$  from a plasmid (PLR2) and R9 represents the amplification from the constant regions of spleen mRNA using primers 11 and 13 (Table 1).

Figure 5 Nucleotide sequences are clones from the cDNA library of the PCR amplified  $V_H$  regions in Lambda ZAP. The N-terminal 110 bases are listed here and the underlined nucleotides represent CDR1 (complementary determining region).

Figure 6 The sequence of the synthetic DNA insert inserted into Lambda ZAP to produce Lambda Zap II  $V_H$  (Panel A) and Lambda Zap  $V_L$  (Panel B) expression vectors. The various features required for this vector to express the  $V_H$  and  $V_L$ -coding DNA homologs include the Shine-Dalgarno ribosome binding site, a leader sequence to direct the expressed protein to the periplasm as described by Mouva et al., J. Biol. Chem., 255:27, 1980, and various restriction enzyme sites used to operatively link the  $V_H$  and  $V_L$  homologs to the expression vector. The  $V_H$  expression-vector sequence also contains a short nucleic acid sequence that codes for amino acids typically found in variable regions heavy chain ( $V_H$  Backbone). This  $V_H$  Backbone is just upstream and in the proper reading as the  $V_H$  DNA homologs that are

operatively linked into the Xho I and Spe I. The  $V_L$  DNA homologs are operatively linked into the  $V_L$  sequence (Panel B) at the Nco I and Spe I restriction enzyme sites and thus the  $V_H$  Backbone region is deleted when the  $V_L$  DNA homologs are operatively linked into the  $V_L$  vector.

Figure 7 The major features of the bacterial expression vector Lambda Zap II  $V_H$  ( $V_H$ -expression vector) are shown. The synthetic DNA sequence from Figure 6 is shown at the top along with the  $T_3$  polymerase promoter from Lambda Zap II. The orientation of the insert in Lambda Zap II is shown. The  $V_H$  DNA homologs are inserted into the Xho I and Spe I restriction enzyme sites. The  $V_H$  DNA are inserted into the Xho I and Spe I site and the read through transcription produces the decapeptide epitope (tag) that is located just 3' of the cloning sites.

Figure 8 The major features of the bacterial expression vector Lambda Zap II  $V_L$  ( $V_L$ -expression vector) are shown. The synthetic sequence shown in Figure 6 is shown at the top along with the  $T_3$  polymerase promoter from Lambda Zap II. The orientation of the insert in Lambda Zap II is shown. The  $V_L$  DNA homologs are inserted into the phagemid that is produced by the in vivo excision protocol described by Short et al., Nucleic Acids Res., 16:7583-7600, 1988. The  $V_L$  DNA homologs are inserted into the Nco I and Spe I cloning sites of the phagemid.

Figure 9 A modified bacterial expression vector Lambda Zap II  $V_L$ II. This vector is constructed by inserting this synthetic DNA sequence,

TGAATTCTAAACTAGTCGCCAAGGAGACAGTCATAATGAA  
TCGAACTTAAGATTTGATCAGCGGTTCTCTGTCAGTATTACTT

5 ATACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACTCGCTG  
TATGGATAACGGATGCCGTCGGCGACCTAACAATAATGAGCGAC

CCCAACCAGCCATGGCCGAGCTCGTCAGTTCTAGAGTTAAGCGGCCG  
GGGTTGGTCGGTACCGGCTCGAGCAGTCAAGATCTCAATTCGCCGGCAGCT

10 into Lambda Zap II that has been digested with the  
restriction enzymes Sac I and Xho I. This sequence  
contains the Shine-Dalgarno sequence (Ribosome  
binding site), the leader sequence to direct the  
expressed protein to the periplasm and the  
15 appropriate nucleic acid sequence to allow the  $V_L$  DNA  
homologs to be operatively linked into the SacI and  
Xba I restriction enzyme sites provided by this  
vector.

20 Figure 10 The sequence of the synthetic DNA  
segment inserted into Lambda Zap II to produce the  
lambda  $V_L$ II-expression vector. The various features  
and restriction endonuclease recognition sites are  
shown.

25 Figure 11 The vectors for expressing  $V_H$  and  
 $V_L$  separately and in combination are shown. The  
various essential components of these vectors are  
shown. The light chain vector or  $V_L$  expression  
vector can be combined with the  $V_H$  expression vector  
to produce a combinatorial vector containing both  $V_H$   
30 and  $V_L$  operatively linked for expression to the same  
promoter.

35 Figure 12 The labelled proteins  
immunoprecipitated from E. coli containing a  $V_H$  and a  
 $V_L$  DNA homolog are shown. In lane 1, the background  
proteins immunoprecipitated from E. coli that do not  
contain a  $V_H$  or  $V_L$  DNA homolog are shown. Lane 2  
contains the  $V_H$  protein immunoprecipitated from E.  
coli containing only a  $V_H$  DNA homolog. In lanes 3

and 4, the commigration of a  $V_H$  protein a  $V_L$  protein immunoprecipitated from E. coli containing both a  $V_H$  and a  $V_L$  DNA homolog is shown. In lane 5 the presence of  $V_H$  protein and  $V_L$  protein expressed from the  $V_H$  and  $V_L$  DNA homologs is demonstrated by the two distinguishable protein species. Lane 5 contains the background proteins immunoprecipitated by anti-E. coli antibodies present in mouse ascites fluid.

Figure 13 The transition state analogue (formula 1) which induces antibodies for hydrolyzing carboxamide substrate (formula 2). The compound of formula 1 containing a glutaryl spacer and a N-hydroxysuccinimide-linker appendage is the form used to couple the hapten (formula 1) to protein carriers KLH and BSA, while the compound of formula 3 is the inhibitor. The phosphoramidate functionality is a mimic of the stereoelectronic features of the transition state for hydrolysis of the amide bond.

Figure 14 PCR amplification of Fd and kappa regions from the spleen mRNA of a mouse immunized with NPN is illustrated. Amplification was performed as described in Example 17 using RNA cDNA hybrids obtained by the reverse transcription of the mRNA with primer specific for amplification of light chain sequences (Table 2) or heavy chain sequences (Table 1). Lanes F1-F8 represent the product of heavy chain amplification reactions with one of each of the eight 5' primers (primers 2-9, Table 1) and the unique 3' primer (primer 15, Table 2). Light chain (k) amplifications with the 5' primers (primers 3-6, and 12, respectively, Table 2) and the appropriate 3' primer (primer 13, Table 2) are shown in lanes F9-F13. A band of 700 bps is seen in all lanes indicating the successful amplification of Fd and k regions.



Figure 15 The screening of phage libraries for antigen binding is depicted according to Example 17C. Duplicate plaque lifts of Fab (filters A,B), heavy chain (filters E,F) and light chain (filters G,H) expression libraries were screened against  $^{125}\text{I}$ -labelled BSA conjugated with NPN at a density of approximately 30,000 plaques per plate. Filters C and D illustrate the duplicate secondary screening of a cored positive from a primary filter A (arrows) as discussed in the text.

Screening employed standard plaque lift methods. XL1 Blue cells infected with phage were incubated on 150mm plates for 4h at  $37^\circ$ , protein expression induced by overlay with nitrocellulose filters soaked in 10mM isopropyl thiogalactoside (IPTG) and the plates incubated at  $25^\circ$  for 8h. Duplicate filters were obtained during a second incubation employing the same conditions. Filters were then blocked in a solution of 1% BSA in PBS for 1h before incubation with rocking at  $25^\circ$  for 1h with a solution of  $^{125}\text{I}$ -labelled BSA conjugated to NPN ( $2 \times 10^6$  cpm  $\text{ml}^{-1}$ ; BSA concentration at 0.1 M; approximately 15 NPN per BSA molecule) in 1% BSA/PBS. Background was reduced by pre-centrifugation of stock radiolabelled BSA solution at 100,000 g for 15 min and pre-incubation of solutions with plaque lifts from plates containing bacteria infected with a phage having no insert. After labeling, filters were washed repeatedly with PBS/0.05% Tween 20 before development of autoradiographs overnight.

Figure 16 The specificity of antigen binding as shown by competitive inhibition is illustrated according to Example 17C. Filter lifts from positive plaques were exposed to  $^{125}\text{I}$ -BSA-NPN in the presence of increasing concentrations of the

inhibitor NPN.

In this study a number of phages correlated with NPN binding as in Figure 15 were spotted (about 100 particles per spot) directly onto a bacterial lawns. The plate was then overlaid with an IPTG-soaked filter and incubated for 19h at 25°. The filter were then blocked in 1% BSA in PBS prior to incubation in <sup>125</sup>I-BSA-NPN as described previously in Figure 15 except with the inclusion of varying amounts of NPN in the labeling solution. Other conditions and procedures were as in Figure 15. The results for a phage of moderate affinity are shown in duplicate in the figure. Similar results were obtained for four other phages with some differences in the effective inhibitor concentration ranges.

Figure 17 The characterization of an antigen binding protein is illustrated according to Example 17D. The concentrated partially purified bacterial supernate of an NPN-binding clone was separated by gel filtration and aliquots from each fraction applied to microtitre plates coated with BSA-NPN. Addition of either anti-decapeptide (---) or anti-kappa chain (—) antibodies conjugated with alkaline phosphatase was followed by color development. The arrow indicates the position of elution of a known Fab fragment. The results show that antigen binding is a property of 50 kD protein containing both heavy and light chains.

Single plaques of two NPN-positive clones (Figure 15) were picked and the plasmid containing the heavy and light chain inserts excised. 500 ml cultures in L-broth were inoculated with 3 ml of a saturated culture containing the excised plasmids and incubated for 4h at 37°. Proteins synthesis was induced by the addition of IPTG to a final

concentration of 1mM and the cultures incubated for 10h at 25°. 200 ml of cells supernate were concentrated to 2 ml and applied to a TSK-G4000 column. 50 ul aliquots from the eluted fractions were assayed by ELISA.

For ELISA analysis, microtitre plates were coated with BSA-NPN at 1 ug/ml, 50 ul samples mixed with 50 ul PBS-Tween 20 (0.05%)-BSA (0.1%) added and the plates incubated for 2h at 25°. After washing with PBS-Tween 20-BSA, 50 ul of appropriate concentrations of a rabbit anti-decapeptide antibody (20) and a goat anti-mouse kappa light chain (Southern Biotech) antibody conjugated with alkaline phosphatase were added and incubated for 2h at 25°. After further washing, 50 ul of p-nitrophenyl phosphate (1 mg/ml in 0.1M tris pH 9.5 containing 50 mM MgCl2) were added and the plates incubated for 15-30 min before reading the OD at 405 nm.

## Detailed Description of the Invention

### A. Definitions

Nucleotide: a monomeric unit of DNA or RNA consisting of a sugar moiety (pentose), a phosphate, and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of the pentose) and that combination of base and sugar is a nucleoside. When the nucleoside contains a phosphate group bonded to the 3' or 5' position of the pentose it is referred to as a nucleotide.

Base Pair (bp): a partnership of adenine (A) with thymine (T), or of cytosine (C) with guanine (G) in a double stranded DNA molecule. In RNA, uracil (U) is substituted for thymine.

Nucleic Acid: a polymer of nucleotides, either single or double stranded.

Gene: a nucleic acid whose nucleotide sequence codes for a RNA or polypeptide. A gene can be either RNA or DNA.

Complementary Bases: nucleotides that normally pair up when DNA or RNA adopts a double stranded configuration.

Complementary Nucleotide Sequence: a sequence of nucleotides in a single-stranded molecule of DNA or RNA that is sufficiently complementary to that on another single strand to specifically hybridize to it with consequent hydrogen bonding.

Conserved: a nucleotide sequence is conserved with respect to a preselected (reference) sequence if it non-randomly hybridizes to an exact complement of the preselected sequence.

Hybridization: the pairing of substantially complementary nucleotide sequences (strands of nucleic acid) to form a duplex or heteroduplex by the establishment of hydrogen bonds between complementary base pairs. It is a specific, i.e. non-random, interaction between two complementary polynucleotide that can be competitively inhibited.

Nucleotide Analog: a purine or pyrimidine nucleotide that differs structurally from a, T, G, C, or U, but is sufficiently similar to substitute for the normal nucleotide in a nucleic acid molecule.

DNA Homolog: is a nucleic acid having a preselected conserved nucleotide sequence and a sequence coding for a receptor capable of binding a preselected ligand

## B. Methods

The present invention provides a novel method for tapping the immunological repertoire by isolating from  $V_H$ -coding and  $V_L$ -coding gene repertoires genes coding for an antibody receptor heterodimeric capable of binding a preselected ligand. Generally, the method combines the following elements:

1. Isolating nucleic acids containing a substantial portion of the immunological repertoire.

2. Preparing polynucleotide primers for cloning polynucleotide segments containing immunoglobulin  $V_H$  and  $V_L$  region genes.

3. Preparing a gene library containing a plurality of different  $V_H$  and  $V_L$  genes from the repertoire.

4. Expressing the  $V_H$  and  $V_L$  polypeptides in a suitable host, including prokaryotic and eukaryotic hosts, either separately or in the same cell, and either on the same or different expression vectors.

5. Screening the expressed polypeptides for the preselected activity, and segregating a  $V_H$ - and  $V_L$ -coding gene combination identified by the screening process.

An antibody produced by the present invention assumes a conformation having a binding site specific for as evidenced by its ability to be competitively inhibited, a preselected or predetermined ligand such as an antigen, enzymatic substrate and the like. In one embodiment, an antibody of this invention forms an antigen binding site which specifically binds to a preselected antigen to form an immunoreaction product (complex)

having a sufficiently strong binding between the antigen and the binding site for the immunoreaction product to be isolated. The antibody typically has an affinity or avidity is generally greater than  $10^5$ - $M^{-1}$  more usually greater than  $10^6$  and preferably greater than  $10^8 M^{-1}$ .

In another embodiment, an antibody of the subject invention binds a substrate and catalyzes the formation of a product from the substrate. While the topology of the ligand binding site of a catalytic antibody is probably more important for its preselected activity than its affinity (association constant or  $pK_a$ ) for the substrate, the subject catalytic antibodies have an association constant for the preselected substrate generally greater than  $10^3 M^{-1}$ , more usually greater than  $10^5 M^{-1}$  or  $10^6 M^{-1}$  and preferably greater than  $10^7 M^{-1}$ .

Preferably the antibody produced by the subject invention is heterodimeric and is therefore normally comprised of two different polypeptide chains, which together assume a conformation having a binding affinity, or association constant for the preselected antibody that is different, preferably higher, than the affinity or association constant of either of the polypeptides alone, i.e., as monomers. One or both of the different polypeptide chains is derived from the variable region of the light and heavy chains of an immunoglobulin. Typically, polypeptides comprising the light ( $V_L$ ) and heavy ( $V_H$ ) variable regions are employed together for binding the preselected antibody.

A  $V_H$  or  $V_L$  produced by the subject invention can be active in monomeric as well as multimeric forms, either homomeric or heteromeric, preferably heterodimeric. A  $V_H$  and  $V_L$  ligand binding

polypeptide produced by the present invention can be advantageously combined in a heterodimer (antibody molecule) to modulate the activity of either or to produce an activity unique to the heterodimer. The individual ligand binding polypeptides will be referred to as  $V_H$  and  $V_L$  and the heterodimer will be referred to as an antibody molecule.

However, it should be understood that a  $V_H$  binding polypeptide may contain in addition to the  $V_H$ , substantially all or a portion of the heavy chain constant region. A  $V_L$  binding polypeptide may contain, in addition to the  $V_L$ , substantially all or a portion of the light chain constant region. A heterodimer comprised of a  $V_H$  binding polypeptide containing a portion of the heavy chain constant region and a  $V_L$  binding containing substantially all of the light chain constant region is termed a Fab fragment. The production of Fab can be advantageous in some situations because the additional constant region sequences contained in a Fab as compared to a  $F_v$  could stabilize the  $V_H$  and  $V_L$  interaction. Such stabilization could cause the Fab to have higher affinity for antigen. In addition the Fab is more commonly used in the art and thus there are more commercial antibodies available to specifically recognize a Fab.

The individual  $V_H$  and  $V_L$  polypeptides will generally have fewer than 125 amino acid residues, more usually fewer than about 120 amino acid residues, while normally having greater than 60 amino acid residues, usually greater than about 95 amino acid residues, more usually greater than about 100 amino acid residues. Preferably, the  $V_H$  will be from about 110 to about 125 amino acid residues in length while  $V_L$  will be from about 95 to about 115 amino

acid residues in length.

The amino acid residue sequences will vary widely, depending upon the particular idio-  
type involved. Usually, there will be at least two  
5 cysteines separated by from about 60 to 75 amino acid  
residues and joined by a disulfide bond. The  
polypeptides produced by the subject invention will  
normally be substantial copies of idotypes of the  
variable regions of the heavy and/or light chains of  
10 immunoglobulins, but in some situations a polypeptide  
may contain random mutations in amino acid residue  
sequences in order to advantageously improve the  
desired activity.

In some situations, it is desirable to  
15 provide for covalent cross linking of the  $V_H$  and  $V_L$   
polypeptides, which can be accomplished by providing  
cysteine residues at the carboxyl termini. The  
polypeptide will normally be prepared free of the  
immunoglobulin constant regions, however a small  
20 portion of the J region may be included as a result  
of the advantageous selection of DNA synthesis  
primers. The D region will normally be included in  
the transcript of the  $V_H$ .

In other situations, it is desirable to  
25 provide a peptide linker to connect the  $V_L$  and the  $V_H$   
to form a single-chain antigen-binding protein  
comprised of a  $V_H$  and a  $V_L$ . This single-chain  
antigen-binding protein would be synthesized as a  
single protein chain. Such single-chain antigen-  
30 binding proteins have been described by Bird et al.,  
Science, 242:423-426 (1988). The design of suitable  
peptide linker regions is described in U.S. Patent  
No. 4,704,692 by Robert Landner.

Such a peptide linker could be designed as  
35 part of the nucleic acid sequences contained in the



expression vector. The nucleic acid sequences coding for the peptide linker would be between the  $V_H$  and  $V_L$  DNA homologs and the restriction endonuclease sites used to operatively link the  $V_H$  and  $V_L$  DNA homologs to the expression vector.

Such a peptide linker could also be coded for nucleic acid sequences that are part of the polynucleotide primers used to prepare the various gene libraries. The nucleic acid sequence coding for the peptide linker can be made up of nucleic acids attached to one of the primers or the nucleic acid sequence coding for the peptide linker may be derived from nucleic acid sequences that are attached to several polynucleotide primers used to create the gene libraries.

Typically the C terminus region of the  $V_H$  and  $V_L$  polypeptides will have a greater variety of the sequences than the N terminus and, based on the present strategy, can be further modified to permit a variation of the normally occurring  $V_H$  and  $V_L$  chains. A synthetic polynucleotide can be employed to vary one or more amino in an hypervariable region.

#### 1. Isolation Of The Repertoire

To prepare a composition of nucleic acids containing a substantial portion of the immunological gene repertoire, a source of genes coding for the  $V_H$  and/or  $V_L$  polypeptides is required. Preferably the source will be a heterogeneous population of antibody producing cells, i.e. B lymphocytes (B cells), preferably rearranged B cells such as those found in the circulation or spleen of a vertebrate.

(Rearranged B cells are those in which immunoglobulin gene translocation, i.e., rearrangement, has occurred as evidenced by the presence in the cell of mRNA with

the immunoglobulin gene V, D and J region transcripts adjacently located thereon.)

5 In some cases, it is desirable to bias the repertoire for a preselected activity, such as by using as a source of nucleic acid cells (source cells) from vertebrates in any one of various stages of age, health and immune response. For example, repeated immunization of a healthy animal prior to collecting rearranged B cells results in obtaining a repertoire enriched for genetic material producing a  
10 ligand binding polypeptide of high affinity. Conversely, collecting rearranged B cells from a healthy animal whose immune system has not been recently challenged results in producing a repertoire that is not biased towards the production of high  
15 affinity  $V_H$  and/or  $V_L$  polypeptides.

It should be noted the greater the genetic heterogeneity of the population of cells for which the nucleic acids are obtained, the greater the  
20 diversity of the immunological repertoire that will be made available for screening according to the method of the present invention. Thus, cells from different individuals, particularly those having an immunologically significant age difference, and cells  
25 from individuals of different strains, races or species can be advantageously combined to increase the heterogeneity of the repertoire.

Thus, in one preferred embodiment, the source cells are obtained from a vertebrate,  
30 preferably a mammal, which has been immunized or partially immunized with an antigenic ligand (antigen) against which activity is sought, i.e., a preselected antigen. The immunization can be carried out conventionally. Antibody titer in the animal can  
35 be monitored to determine the stage of immunization

desired, which stage corresponds to the amount of enrichment or biasing of the repertoire desired. Partially immunized animals typically receive only one immunization and cells are collected therefrom shortly after a response is detected. Fully immunized animals display a peak titer, which is achieved with one or more repeated injections of the antigen into the host mammal, normally at 2 to 3 week intervals. Usually three to five days after the last challenge, the spleen is removed and the genetic repertoire of the spleenocytes, about 90% of which are rearranged B cells, is isolated using standard procedures. See, Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley & Sons, NY.

Nucleic acids coding for  $V_H$  and  $V_L$  polypeptides can be derived from cells producing IgA, IgD, IgE, IgG or IgM, most preferably from IgM and IgG, producing cells.

Methods for preparing fragments of genomic DNA from which immunoglobulin variable region genes can be cloned as a diverse population are well known in the art. See for example Herrmann et al., Methods In Enzymol., 152:180-183, (1987); Frischauf, Methods In Enzymol., 152:183-190 (1987); Frischauf, Methods In Enzymol., 152:190-199 (1987); and DiLella et al., Methods In Enzymol., 152:199-212 (1987). (The teachings of the references cited herein are hereby incorporated by reference.)

The desired gene repertoire can be isolated from either genomic material containing the gene expressing the variable region or the messenger RNA (mRNA) which represents a transcript of the variable region. The difficulty in using the genomic DNA from other than non-rearranged B lymphocytes is in juxtaposing the sequences coding for the variable

region, where the sequences are separated by introns. The DNA fragment(s) containing the proper exons must be isolated, the introns excised, and the exons then spliced in the proper order and in the proper orientation. For the most part, this will be difficult, so that the alternative technique employing rearranged B cells will be the method of choice because the C D and J immunoglobulin gene regions have translocated to become adjacent, so that the sequence is continuous (free of introns) for the entire variable regions.

Where mRNA is utilized the cells will be lysed under RNase inhibiting conditions. In one embodiment, the first step is to isolate the total cellular mRNA by hybridization to an oligo-dT cellulose column. The presence of mRNAs coding for the heavy and/or light chain polypeptides can then be assayed by hybridization with DNA single strands of the appropriate genes. Conveniently, the sequences coding for the constant portion of the  $V_H$  and  $V_L$  can be used as polynucleotide probes, which sequences can be obtained from available sources. See for example, Early and Hood, Genetic Engineering, Setlow and Hollaender, eds., Vol. 3, Plenum Publishing Corporation, New York, (1981), pages 157-188; and Kabat et al., Sequences of Immunological Interest, National Institutes of Health, Bethesda, MD, (1987).

In preferred embodiments, the preparation containing the total cellular mRNA is first enriched for the presence of  $V_H$  and/or  $V_L$  coding mRNA. Enrichment is typically accomplished by subjecting the total mRNA preparation or partially purified mRNA product thereof to a primer extension reaction employing a polynucleotide synthesis primer of the present invention.

## 2. Preparation Of Polynucleotide Primers

The term "polynucleotide" as used herein in reference to primers, probes and nucleic acid fragments or segments to be synthesized by primer extension is defined as a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, preferably more than 3. Its exact size will depend on many factors, which in turn depends on the ultimate conditions of use.

The term "primer" as used herein refers to a polynucleotide whether purified from a nucleic acid restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, i.e., in the presence of nucleotides and an agent for polymerization such as DNA polymerase, reverse transcriptase and the like, and at a suitable temperature and pH. The primer is preferably single stranded for maximum efficiency, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is a polydeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the agents for polymerization. The exact lengths of the primers will depend on many factors, including temperature and the source of primer. For example, depending on the complexity of the target sequence, a polynucleotide primer typically contains 15 to 25 or more nucleotides, although it can contain fewer nucleotides. Short primer molecules generally

require cooler temperatures to form sufficiently stable hybrid complexes with template.

5 The primers used herein are selected to be "substantially" complementary to the different strands of each specific sequence to be synthesized or amplified. This means that the primer must be sufficiently complementary to nonrandomly hybridize with its respective template strand. Therefore, the primer sequence may not reflect the exact sequence of 10 the template. For example, a non-complementary nucleotide fragment can be attached to the 5' end of the primer, with the remainder of the primer sequence being substantially complementary to the strand. Such noncomplementary fragments typically code for an endonuclease restriction site. Alternatively, 15 noncomplementary bases or longer sequences can be interspersed into the primer, provided the primer sequence has sufficient complementarity with the sequence of the strand to be synthesized or amplified to non-randomly hybridize therewith and thereby form an extension product under polynucleotide synthesizing conditions. 20

25 The polynucleotide primers can be prepared using any suitable method, such as, for example, the phosphotriester or phosphodiester methods see Narang et al., Meth. Enzymol., 68:90, (1979); U.S. Patent No. 4,356,270; and Brown et al., Meth. Enzymol., 68:109, (1979).

30 The choice of a primer's nucleotide sequence depends on factors such as the distance on the nucleic acid from the region coding for the desired receptor, its hybridization site on the nucleic acid relative to any second primer to be used, the number of genes in the repertoire it is to hybridize to, and 35 the like.

For example, to produce  $V_H$ -coding DNA homologs by primer extension, the nucleotide sequence of a primer is selected to hybridize with a plurality of immunoglobulin heavy chain genes at a site substantially adjacent to the  $V_H$ -coding region so that a nucleotide sequence coding for a functional (capable of binding) polypeptide is obtained. To hybridize to a plurality of different  $V_H$ -coding nucleic acid strands, the primer must be a substantial complement of a nucleotide sequence conserved among the different strands. Such sites include nucleotide sequences in the constant region, any of the variable region framework regions, preferably the third framework region, leader region, promoter region, J region and the like.

If the  $V_H$ -coding and  $V_L$ -coding DNA homologs are to be produced by polymerase chain reaction (PCR) amplification, two primers must be used for each coding strand of nucleic acid to be amplified. The first primer becomes part of the nonsense (minus or complimentary) strand and hybridizes to a nucleotide sequence conserved among  $V_H$  (plus) strands within the repertoire. To produce  $V_H$  coding DNA homologs, first primers are therefore chosen to hybridize to (i.e. be complementary to) conserved regions within the J region, CH1 region, hinge region, CH2 region, or CH3 region of immunoglobulin genes and the like. To produce a  $V_L$  coding DNA homolog, first primers are chosen to hybridize with (i.e. be complementary to) a conserved region within the J region or constant region of immunoglobulin light chain genes and the like. Second primers become part of the coding (plus) strand and hybridize to a nucleotide sequence conserved among minus strands. To produce the  $V_H$ -coding DNA homologs, second primers are therefore

chosen to hybridize with a conserved nucleotide sequence at the 5' end of the V<sub>H</sub>-coding immunoglobulin gene such as in that area coding for the leader or first framework region. It should be noted that in the amplification of both V<sub>H</sub>- and V<sub>L</sub>-coding DNA homologs the conserved 5' nucleotide sequence of the second primer can be complementary to a sequence exogenously added using terminal deoxynucleotidyl transferase as described by Loh et al., Sci. Vol 243:217-220 (1989). One or both of the first and second primers can contain a nucleotide sequence defining an endonuclease recognition site. The site can be heterologous to the immunoglobulin gene being amplified and typically appears at or near the 5' end of the primer.

### 3. Preparing a Gene Library

The strategy used for cloning, i.e., substantially reproducing, the V<sub>H</sub> and/or V<sub>L</sub> genes contained within the isolated repertoire will depend, as is well known in the art, on the type, complexity, and purity of the nucleic acids making up the repertoire. Other factors include whether or not the genes are to be amplified and/or mutagenized.

In one strategy, the object is to clone the V<sub>H</sub>- and/or V<sub>L</sub>-coding genes from a repertoire comprised of polynucleotide coding strands, such as mRNA and/or the sense strand of genomic DNA. If the repertoire is in the form of double stranded genomic DNA, it is usually first denatured, typically by melting, into single strands. The repertoire is subjected to a first primary extension reaction by treating (contacting) the repertoire with a first polynucleotide synthesis primer having a preselected nucleotide sequence. The first primer is capable of



initiating the first primer extension reaction by hybridizing to a nucleotide sequence, preferably at least about 10 nucleotides in length and more preferably at least about 20 nucleotides in length, conserved within the repertoire. The first primer is sometimes referred to herein as the "sense primer" because it hybridizes to the coding or sense strand of a nucleic acid. In addition, the second primer is sometimes referred to herein as the "anti-sense primer" because it hybridizes to a non-coding or anti-sense strand of a nucleic acid, i.e., a strand complementary to a coding strand.

The first primer extension is performed by mixing the first primer, preferably a predetermined amount thereof, with the nucleic acids of the repertoire, preferably a predetermined amount thereof, to form a first primer extension reaction admixture. The admixture is maintained under polynucleotide synthesizing conditions for a time period, which is typically predetermined, sufficient for the formation of a first primer extension reaction product, thereby producing a plurality of different  $V_H$ -coding DNA homolog complements. The complements are then subjected to a second primer extension reaction by treating them with a second polynucleotide synthesis primer having a preselected nucleotide sequence. The second primer is capable of initiating the second reaction by hybridizing to a nucleotide sequence, preferably at least about 10 nucleotides in length and more preferably at least about 20 nucleotides in length, conserved among a plurality of different  $V_H$ -coding gene complements such as those, for example, produced by the first primer extension reaction. This is accomplished by mixing the second primer, preferably a predetermined

amount thereof, with the complement nucleic acids, preferably a predetermined amount thereof, to form a second primer extension reaction admixture. The admixture is maintained under polynucleotide synthesizing conditions for a time period, which is typically predetermined, sufficient for the formation of a first primer extension reaction product, thereby producing a gene library containing a plurality of different  $V_H$ -and/or  $V_L$ -coding DNA homologs.

A plurality of first primer and/or a plurality of second primers can be used in each amplification, or an individual pair of first and second primers can be used. In any case, the amplification products of amplifications using the same or different combinations of first and second primers can be combined to increase the diversity of the gene library.

In another strategy, the object is to clone the  $V_H$ - and/or  $V_L$ -coding genes from a repertoire by providing a polynucleotide complement of the repertoire, such as the anti-sense strand of genomic dsDNA or the polynucleotide produced by subjecting mRNA to a reverse transcriptase reaction. Methods for producing such complements are well known in the art. The complement is subjected to a primer extension reaction similar to the above-described second primer extension reaction, i.e., a primer extension reaction using a polynucleotide synthesis primer capable of hybridizing to a nucleotide sequence conserved among a plurality of different  $V_H$ -coding gene complements.

The primer extension reaction is performed using any suitable method. Generally it occurs in a buffered aqueous solution, preferably at a pH of 7-9, most preferably about 8. Preferably, a molar excess

(for genomic nucleic acid, usually about  $10^6:1$  primer:template) of the primer is admixed to the buffer containing the template strand. A large molar excess is preferred to improve the efficiency of the process.

The deoxyribonucleotide triphosphates dATP, dCTP, dGTP, and dTTP are also admixed to the primer extension (polynucleotide synthesis) reaction admixture in adequate amounts and the resulting solution is heated to about 90C - 100C for about 1 to 10 minutes, preferably from 1 to 4 minutes. After this heating period the solution is allowed to cool to room temperature, which is preferable for primer hybridization. To the cooled mixture is added an appropriate agent for inducing or catalyzing the primer extension reaction, and the reaction is allowed to occur under conditions known in the art. The synthesis reaction may occur at from room temperature up to a temperature above which the inducing agent no longer functions efficiently. Thus, for example, if DNA polymerase is used as inducing agent, the temperature is generally no greater than about 40C.

The inducing agent may be any compound or system which will function to accomplish the synthesis of primer extension products, including enzymes. Suitable enzymes for this purpose include, for example, E. coli, DNA polymerase I, Klenow fragment of E. coli DNA polymerase I, T4 DNA polymerase, other available DNA polymerases, reverse transcriptase, and other enzymes, including heat-stable enzymes, which will facilitate combination of the nucleotides in the proper manner to form the primer extension products which are complementary to each nucleic acid strand. Generally, the synthesis

will be initiated at the 3' end of each primer and proceed in the 5' direction along the template strand, until synthesis terminates, producing molecules of different lengths. There may be  
5 inducing agents, however, which initiate synthesis at the 5' end and proceed in the above direction, using the same process as described above.

The newly synthesized strand and its complementary nucleic acid strand form a double-  
10 stranded molecule which can be used in the succeeding steps of the process.

The first and/or second primer extension reaction discussed above can advantageously be used to incorporate into the receptor a preselected epitope useful in immunologically detecting and/or  
15 isolating a receptor. This is accomplished by utilizing a first and/or second polynucleotide synthesis primer or expression vector to incorporate a predetermined amino acid residue sequence into the amino acid residue sequence of the receptor.  
20

After producing  $V_H$ - and/or  $V_L$ -coding DNA homologs for a plurality of different  $V_H$ - and/or  $V_L$ -coding genes within the repertoire, the homologs are typically amplified. While the  $V_H$  and/or  $V_L$ -coding  
25 DNA homologs can be amplified by classic techniques such as incorporation into an autonomously replicating vector, it is preferred to first amplify the DNA homologs by subjecting them to a polymerase chain reaction (PCR) prior to inserting them into a  
30 vector. In fact, in preferred strategies, the first and/or second primer extension reactions used to produce the gene library are the first and second primer extension reactions in a polymerase chain reaction.

PCR is typically carried out by cycling i.e., simultaneously performing in one admixture, the above described first and second primer extension reactions, each cycle comprising polynucleotide synthesis followed by denaturation of the double stranded polynucleotides formed. Methods and systems for amplifying a DNA homolog are described in U.S. Patents No. 4,683,195 and No. 4,683,202, both to Mullis et al.

In preferred embodiments only one pair of first and second primers is used per amplification reaction. The amplification reaction products obtained from a plurality of different amplifications, each using a plurality of different primer pairs, are then combined.

However, the present invention also contemplates DNA homolog production via co-amplification (using two pairs of primers), and multiplex amplification (using up to about 8, 9 or 10 primer pairs).

The  $V_H$ - and  $V_L$ -coding DNA homologs produced by PCR amplification are typically in double-stranded form and have contiguous or adjacent to each of their termini a nucleotide sequence defining an endonuclease restriction site. Digestion of the  $V_H$ - and  $V_L$ -coding DNA homologs having restriction sites at or near their termini with one or more appropriate endonucleases results in the production of homologs having coheasive termini of predetermined specificity.

In preferred embodiments, the PCR process is used not only to amplify the  $V_H$ - and/or  $V_L$ -coding DNA homologs of the library, but also to induce mutations within the library and thereby provide a library having a greater heterogeneity. First, it should be

noted that the PCR processes itself is inherently mutagenic due to a variety of factors well known in the art. Second, in addition to the mutation inducing variations described in the above referenced U.S. Patent No. 4,683,195, other mutation inducing PCR variations can be employed. For example, the PCR reaction admixture, i.e., the combined first and second primer extension reaction admixtures, can be formed with different amounts of one or more of the nucleotides to be incorporated into the extension product. Under such conditions, the PCR reaction proceeds to produce nucleotide substitutions within the extension product as a result of the scarcity of a particular base. Similarly, approximately equal molar amounts of the nucleotides can be incorporated into the initial PCR reaction admixture in an amount to efficiently perform X number of cycles, and then cycling the admixture through a number of cycles in excess of X, such as, for instance, 2X. Alternatively, mutations can be induced during the PCR reaction by incorporating into the reaction admixture nucleotide derivatives such as inosine, not normally found in the nucleic acids of the repertoire being amplified. During subsequent in vivo amplification, the nucleotide derivative will be replaced with a substitute nucleotide thereby inducing a point mutation.

#### 4. Expressing the V<sub>H</sub> and/or V<sub>L</sub> DNA Homologs.

The V<sub>H</sub> and/or V<sub>L</sub>-coding DNA homologs contained within the library produced by the above-described method can be operatively linked to a vector for amplification and/or expression.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting between different genetic environments another nucleic acid to which it has been operatively linked. One type of preferred vector is an episome, i.e., a nucleic acid molecule capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors".

The choice of vector to which a  $V_H$ - and/or  $V_L$ -coding DNA homolog is operatively linked depends directly, as is well known in the art, on the functional properties desired, e.g., replication or protein expression, and the host cell to be transformed, these being limitations inherent in the art of constructing recombinant DNA molecules. In preferred embodiments, the vector utilized includes a procaryotic replicon i.e., a DNA sequence having the ability to direct autonomous replication and maintenance of the recombinant DNA molecule extra chromosomally in a procaryotic host cell, such as a bacterial host cell, transformed therewith. Such replicons are well known in the art. In addition, those embodiments that include a procaryotic replicon also include a gene whose expression confers a selective advantage, such as drug resistance, to a bacterial host transformed therewith. Typical bacterial drug resistance genes are those that confer resistance to ampicillin or tetracycline.

Those vectors that include a procaryotic replicon can also include a procaryotic promoter capable of directing the expression (transcription

and translation) of the  $V_H$ - and/or  $V_L$ -coding homologs in a bacterial host cell, such as E. coli transformed therewith. A promoter is an expression control element formed by a DNA sequence that permits binding of RNA polymerase and transcription to occur.

Promoter sequences compatible with bacterial hosts are typically provided in plasmid vectors containing convenience restriction sites for insertion of a DNA segment of the present invention. Typical of such vector plasmids are pUC8, pUC9, pBR322, and pBR329 available from BioRad Laboratories, (Richmond, CA) and pPL and pKK223 available from Pharmacia, (Piscataway, NJ).

Expression vectors compatible with eucaryotic cells, preferably those compatible with vertebrate cells, can also be used. Eucaryotic cell expression vectors are well known in the art and are available from several commercial sources.

Typically, such vectors are provided containing convenient restriction sites for insertion of the desired DNA homologue. Typical of such vectors are pSVL and pKSV-10 (Pharmacia), pBPV-1/PML2d (International Biotechnologies, Inc.), and pTDT1 (ATCC, No. 31255).

In preferred embodiments, the eucaryotic cell expression vectors used include a selection marker that is effective in an eucaryotic cell, preferably a drug resistant selection marker. A preferred drug resistance marker is the gene whose expression results in neomycin resistance, i.e., the neomycin phosphotransferase (neo) gene. Southern et al., J. Mol. Appl. Genet., 1:327-341 (1982).

The use of retroviral expression vectors to express the genes of the  $V_H$  and/or  $V_L$ -coding DNA homologs is also contemplated. As used herein, the



term "retroviral expression vector" refers to a DNA molecule that includes a promoter sequences derived from the long terminal repeat (LTR) region of a retrovirus genome.

5 In preferred embodiments, the expression vector is typically a retroviral expression vector that is preferably replication-incompetent in eucaryotic cells. The construction and use of retroviral vectors has been described by Sorge et  
10 al., Mol. Cel. Biol., 4:1730-1737 (1984).

A variety of methods have been developed to operatively link DNA to vectors via complementary cohesive termini. For instance, complementary cohesive termini can be engineered into the  $V_H$ -  
15 and/or  $V_L$ -coding DNA homologs during the primer extension reaction by use of an appropriately designed polynucleotide synthesis primer, as previously discussed. The vector, and DNA homolog if necessary, is cleaved with a restriction endonuclease to produce termini complementary to those of the DNA  
20 homolog. The complementary cohesive termini of the vector and the DNA homolog are then operatively linked (ligated) to produce a unitary double stranded DNA molecule.

25 In preferred embodiments, the  $V_H$ -coding and  $V_L$ -coding DNA homologs of diverse libraries are randomly combined in vitro for polycistronic expression from individual vectors. That is, a diverse population of double stranded DNA expression  
30 vectors is produced wherein each vector expresses, under the control of a single promoter, one  $V_H$ -coding DNA homolog and one  $V_L$ -coding DNA homolog, the diversity of the population being the result of different  $V_H$ - and  $V_L$ -coding DNA homolog combinations.

Random combination in vitro can be accomplished using two expression vectors distinguished from one another by the location on each of a restriction site common to both. Preferably the vectors are linear double stranded DNA, such as a Lambda Zap derived vector as described herein. In the first vector, the site is located between a promoter and a polylinker, i.e., 5' terminal (upstream relative to the direction of expression) to the polylinker but 3' terminal (downstream relative to the direction of expression). In the second vector, the polylinker is located between a promoter and the restriction site, i.e., the restriction site is located 3' terminal to the polylinker, and the polylinker is located 3' terminal to the promoter.

In preferred embodiments, each of the vectors defines a nucleotide sequence coding for a ribosome binding and a leader, the sequence being located between the promoter and the polylinker, but downstream (3' terminal) from the shared restriction site if that site is between the promoter and polylinker. Also preferred are vectors containing a stop codon downstream from the polylinker, but upstream from any shared restriction site if that site is downstream from the polylinker. The first and/or second vector can also define a nucleotide sequence coding for a peptide tag. The tag sequence is typically located downstream from the polylinker but upstream from any stop codon that may be present.

In preferred embodiments, the vectors contain selectable markers such that the presence of a portion of that vector, i.e. a particular lambda arm, can be selected for or selected against. Typical selectable markers are well known to those

skilled in the art. Examples of such markers are antibiotic resistance genes, genetically selectable markers, mutation suppressors such as amber suppressors and the like. The selectable markers are typically located upstream of the promoter and/or downstream of the second restriction site. In preferred embodiments, one selectable marker is located upstream of the promoter on the first vector containing the  $V_H$ -coding DNA homologs. A second selectable marker is located downstream of the second restriction site on the vector containing the  $V_L$ -coding DNA homologs. This second selectable marker may be the same or different from the first as long as when the  $V_H$ -coding vectors and the  $V_L$ -coding vectors are randomly combined via the first restriction site the resulting vectors containing both  $V_H$  and  $V_L$  and both selectable markers can be selected.

Typically the polylinker is a nucleotide sequence that defines one or more, preferably at least two, restriction sites, each unique to the vector and preferably not shared by the other vector, i.e., if it is on the first vector, it is not on the second vector. The polylinker restriction sites are oriented to permit ligation of  $V_H$ - or  $V_L$ -coding DNA homologs into the vector in same reading frame as any leader, tag or stop codon sequence present.

Random combination is accomplished by ligating  $V_H$ -coding DNA homologs into the first vector, typically at a restriction site or sites within the polylinker. Similarly,  $V_L$ -coding DNA homologs are ligated into the second vector, thereby creating two diverse populations of expression vectors. It does not matter which type of DNA

homolog, i.e.,  $V_H$  or  $V_L$ , is ligated to which vector, but it is preferred, for example, that all  $V_H$ -coding DNA homologs are ligated to either the first or second vector, and all of the  $V_L$ -coding DNA homologs are ligated to the other of the first or second vector. The members of both populations are then cleaved with an endonuclease at the shared restriction site, typically by digesting both populations with the same enzyme. The resulting product is two diverse populations of restriction fragments where the members of one have cohesive termini complementary to the cohesive termini of the members of the other. The restriction fragments of the two populations are randomly ligated to one another, i.e., a random, interpopulation ligation is performed, to produce a diverse population of vectors each having a  $V_H$ -coding and  $V_L$ -coding DNA homolog located in the same reading frame and under the control of second vector's promoter. Of course, subsequent recombinations can be effected through cleavage at the shared restriction site, which is typically reformed upon ligation of members from the two populations, followed by subsequent religations.

The resulting construct is then introduced into an appropriate host to provide amplification and/or expression of the  $V_H$ - and/or  $V_L$ -coding DNA homologs, either separately or in combination. When coexpressed within the same organism, either on the same or the different vectors, a functionally active Fv is produced. When the  $V_H$  and  $V_L$  polypeptides are expressed in different organisms, the respective polypeptides are isolated and then combined in an appropriate medium to form a Fv. Cellular hosts into which a  $V_H$ - and/or  $V_L$ -coding DNA homolog-containing construct has been introduced are referred to herein

as having been "transformed" or as "transformants".

The host cell can be either procaryotic or eucaryotic. Bacterial cells are preferred procaryotic host cells and typically are a strain of E. coli such as, for example, the E. coli strain DH5 available from Bethesda Research Laboratories, Inc., Bethesda, MD. Preferred eucaryotic host cells include yeast and mammalian cells, preferably vertebrate cells such as those from a mouse, rat, monkey or human cell line.

Transformation of appropriate cell hosts with a recombinant DNA molecule of the present invention is accomplished by methods that typically depend on the type of vector used. With regard to transformation of procaryotic host cells, see, for example, Cohen et al., Proceedings National Academy of Science, USA, Vol. 69, P. 2110 (1972); and Maniatis et al., Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982). With regard to the transformation of vertebrate cells with retroviral vectors containing rDNAs, see for example, Sorge et al., Mol. Cell. Biol., 4:1730-1737 (1984); Graham et al., Virol., 52:456 (1973); and Wigler et al., Proceedings National Academy of Sciences, USA, Vol. 76, P. 1373-1376 (1979).

##### 5. Screening For Expression of $V_H$ and/or $V_L$ Polypeptides

Successfully transformed cells, i.e., cells containing a  $V_H$ - and/or  $V_L$ -coding DNA homolog operatively linked to a vector, can be identified by any suitable well known technique for detecting the binding of a receptor to a ligand or the presence of a polynucleotide coding for the receptor, preferably

its active site. Preferred screening assays are those where the binding of ligand by the receptor produces a detectable signal, either directly or indirectly. Such signals include, for example, the production of a complex, formation of a catalytic reaction product, the release or uptake of energy, and the like. For example, cells from a population subjected to transformation with a subject rDNA can be cloned to produce monoclonal colonies. Cells from those colonies can be harvested, lysed and their DNA content examined for the presence of the rDNA using a method such as that described by Southern, J. Mol. Biol., 98:503 (1975) or Berent et al., Biotech. 3:208 (1985).

In addition to directly assaying for the presence of a  $V_H$ - and/or  $V_L$ -coding DNA homolog, successful transformation can be confirmed by well known immunological methods, especially when the  $V_H$  and/or  $V_L$  polypeptides produced contain a preselected epitope. For example, samples of cells suspected of being transformed are assayed for the presence of the preselected epitope using an antibody against the epitope.

#### 6. $V_H$ - And/Or $V_L$ -Coding Gene Libraries

The present invention contemplates a gene library, preferably produced by a primer extension reaction or combination of primer extension reactions as described herein, containing at least about  $10^3$ , preferably at least about  $10^4$  and more preferably at least about  $10^5$  different  $V_H$ - and/or  $V_L$ -coding DNA homologs. The homologs are preferably in an isolated form, that is, substantially free of materials such as, for example, primer extension reaction agents and/or substrates, genomic DNA segments, and the

like.

In preferred embodiments, a substantial portion of the homologs present in the library are operatively linked to a vector, preferably  
 5 operatively linked for expression to an expression vector.

Preferably, the homologs are present in a medium suitable for in vitro manipulation, such as water, water containing buffering salts, and the  
 10 like. The medium should be compatible with maintaining the biological activity of the homologs. In addition, the homologs should be present at a concentration sufficient to allow transformation of a host cell compatible therewith at reasonable  
 15 frequencies.

It is further preferred that the homologs be present in compatible host cells transformed therewith.

#### 20 D. Expression Vectors

The present invention also contemplates various expression vectors useful in performing, inter alia, the methods of the present invention. Each of the expression vectors is a novel derivative  
 25 of Lambda Zap.

##### 1. Lambda Zap II

Lambda Zap II is prepared by replacing the Lambda S gene of the vector Lambda Zap with the Lambda S gene from the Lambda gt10 vector, as  
 30 described in Example 6.

##### 2. Lambda Zap II V<sub>H</sub>

Lambda Zap II V<sub>H</sub> is prepared by inserting the synthetic DNA sequences illustrated in Figure 6A into the above-described Lambda Zap II  
 35 vector. The inserted nucleotide sequence

advantageously provides a ribosome binding site (Shine-Dalgarno sequence) to permit proper initiation of mRNA translation into protein, and a leader sequence to efficiently direct the translated protein to the periplasm. The preparation of Lambda Zap II  $V_H$  is described in more detail in Example 9, and its features illustrated in Figures 6A and 7.

### 3. Lambda Zap II $V_L$

Lambda Zap II  $V_L$  is prepared as described in Example 12 by inserting into Lambda Zap II the synthetic DNA sequence illustrated in Figure 6B. Important features of Lambda Zap II  $V_L$  are illustrated in Figure 8.

### 4. Lambda Zap II $V_L$ II

Lambda Zap II  $V_L$  II is prepared as described in Example 11 by inserting into Lambda Zap II the synthetic DNA sequence illustrated in Figure 10.

The above-described vectors are compatible with E. coli hosts, i.e., they can express for secretion into the periplasm proteins coded for by genes to which they have been operatively linked for expression.

## 25 Examples

The following examples are intended to illustrate, but not limit, the scope of the invention.

### 30 1. Polynucleotide Selection

The nucleotide sequences encoding the immunoglobulin protein CDR's are highly variable. However, there are several regions of conserved sequences that flank the  $V_H$  domains. For instance, contain substantially conserved nucleotide sequences,



i.e., sequences that will hybridize to the same primer sequence. Therefore, polynucleotide synthesis (amplification) primers that hybridize to the conserved sequences and incorporate restriction sites into the DNA homolog produced that are suitable for operatively linking the synthesized DNA fragments to a vector were constructed. More specifically, the DNA homologs were inserted into Lambda ZAP II vector (Stratagene Cloning System, San Diego, CA) at the Xho I and EcoR I sites. For amplification of the  $V_H$  domains, the 3' primer (primer 12 in Table 1), was designed to be complementary to the mRNA in the  $J_H$  region. In all cases, the 5' primers (primers 1-10, Table 1) were chosen to be complementary to the first strand cDNA in the conserved N-terminus region (antisense strand). Initially amplification was performed with a mixture of 32 primers (primer 1, Table 1) that were degenerate at five positions. Hybridoma mRNA could be amplified with mixed primers, but initial attempts to amplify mRNA from spleen yielded variable results. Therefore, several alternatives to amplification using the mixed 5' primers were compared.

The first alternative was to construct multiple unique primers, eight of which are shown in Table 1, corresponding to individual members of the mixed primer pool. The individual primers 2-9 of Table 1 were constructed by incorporating either of the two possible nucleotides at three of the five degenerate positions.

The second alternative was to construct a primer containing inosine (primer 10, Table 1) at four of the variable positions based on the published work of Takahashi, et al., Proc. Natl. Acad. Sci. (U.S.A.), 82:1931-1935, (1985) and Ohtsuka et al., J.

Biol. Chem., 260: 2605-2608, (1985). This primer has the advantage that it is not degenerate and, at the same time minimizes the negative effects of mismatches at the unconserved positions as discussed by Martin et al., Nuc. Acids Res., 13:8927 (1985). However, it was not known if the presence of inosine nucleotides would result in incorporation of unwanted sequences in the cloned  $V_H$  regions. Therefore, inosine was not included at the one position that remains in the amplified fragments after the cleavage of the restriction sites. As a result, inosine was not in the cloned insert.

Additional,  $V_H$  amplification primers including the unique 3' primer were designed to be complementary to a portion of the first constant region domain of the gamma 1 heavy chain mRNA (primers 15 and 16, Table 1). These primers will produce DNA homologs containing polynucleotides coding for amino acids from the  $V_H$  and the first constant region domains of the heavy chain. These DNA homologs can therefore be used to produce Fab fragments rather than an  $F_v$ .

As a control for amplification from spleen or hybridoma mRNA, a set of primers hybridizing to a highly conserved region within the constant region IgG, heavy chain gene were constructed. The 5' primer (primer 11, Table 1) is complementary to the cDNA in the  $C_H2$  region whereas the 3' primer (primer 13, Table 1) is complementary to the mRNA in the  $C_H3$  region. It is believed that no mismatches were present between these primers and their templates.

The nucleotide sequences encoding the  $V_L$  CDRs are highly variable. However, there are several regions of conserved sequences that flank the  $V_L$  CDR domains including the  $J_L$ ,  $V_L$  framework regions and  $V_L$

leader/promotor. Therefore, amplification primers that hybridize to the conserved sequences and incorporate restriction sites that allowing cloning the amplified fragments into the pBluescript SK-  
 5 vector cut with Nco I and Spe I were constructed. For amplification of the  $V_L$  CDR domains, the 3' primer (primer 14 in Table 1), was designed to be complementary to the mRNA in the  $J_L$  regions. The 5' primer (primer 15, Table 1) was chosen to be  
 10 complementary to the first strand cDNA in the conserved N-terminus region (antisense strand).

A second set of amplification primers for amplification of the  $V_L$  CDR domains the 5' primers (primers 1-8 in Table 2) were designed to be  
 15 complementary to the first strand cDNA in the conserved N-terminus region. These primers also introduced a Sac I restriction endonuclease site to allow the  $V_L$  DNA homolog to be cloned into the  $V_L$  II-expression vector. The 3'  $V_L$  amplification primer  
 20 (primer 9 in Table 2) was designed to be complementary to the mRNA in the  $J_L$  regions and to introduce the Xba I restriction endonuclease site required to insert the  $V_L$  DNA homolog into the  $V_L$  II-expression vector (Figure a).

25 Additional 3'  $V_L$  amplification primers were designed to hybridize to the constant region of either kappa or lambda mRNA (primers 10 and 11 in Table 2). These primers allow a DNA homolog to be produced containing polynucleotide sequences coding  
 30 for constant region amino acids of either kappa or lambda chain. These primers make it possible to produce an Fab fragment rather than an  $F_v$ .

The primers used for amplification of kappa light chain sequences for construction of Fabs are  
 35 shown at least in Table 2. Amplification with these

primers was performed in 5 separate reactions, each containing one of the 5' primers (primers 3-6, and 12) and one of the 3' primers (primer 13). The remaining 3' primer (primer 9) has been used to construct  $F_v$  fragments. The 5' primers contain a Sac I restriction site and the 3' primers contain a Xba I restriction site.

The primers used for amplification of heavy chain Fd fragments for construction of Fabs are shown at least in Table 1. Amplification was performed in eight separate reactions, each containing one of the 5' primers (primers 2-9) and one of the 3' primers (primer 15). The remaining 5' primers that have been used for amplification in a single reaction are either a degenerate primer (primer 1) or a primer that incorporates inosine at four degenerate positions (primer 10, Table 1, and primers 17 and 18, Table 2). The remaining 3' primer (primer 14, Table 2) has been used to construct  $F_v$  fragments. Many of the 5' primers incorporate a Xho I site, and the 3' primers include a Spe I restriction site.

$V_L$  amplification primers designed to amplify human light chain variable regions of both the lambda and kappa isotypes are also shown in Table 2.

All primers and synthetic polynucleotides used herein and shown on Tables 1-4 were either purchased from Research Genetics in Huntsville, Alabama or synthesized on an Applied Biosystems DNA synthesizer, model 381A, using the manufacturer's instruction.

TABLE 1

5' AGGT(C/G)(C/A)A(G/A)CT(G/T)CTCGAGTC(T/A)GG 3'	degenerate 5' primer for the amplification of variable heavy chain region ( $V_H$ )
5' AGGTCCAGCTGCTCGAGTCTGG 3'	Unique 5' primer for the amplification of $V_H$
5' AGGTCCAGCTGCTCGAGTCAGG 3'	"
5' AGGTCCAGCTTCTCGAGTCTGG 3'	"
5' AGGTCCAGCTTCTCGAGTCAGG 3'	"
5' AGGTCCAACTGCTCGAGTCTGG 3'	"
5' AGGTCCAACTGCTCGAGTCAGG 3'	"
5' AGGTCCAACTTCTCGAGTCTGG 3'	"
5' AGGTCCAACTTCTCGAGTCAGG 3'	"
5' AGGTIIAICTICTCGAGTC(T/A) 3'	5' degenerate primer containing inosine at 4 degenerate positions
5' GCCCAAGGATGTGCTCACC 3'	5' primer for amplification in the $C_H2$ region of mouse IgG1
5' CTATTAGAATTCAACGGTAACAGTGGTGCCTTGGCCCCA 3'	3' primer for amplification of $V_H$
5' CTATTAACTAGTAACGGTAACAGTGGTGCCTTGGCCCCA 3'	3' primer for amplification of $V_H$ using 3' Spe I site
5' CTCAGTATGGTGGTTGTGC 3'	3' primer for amplification in the $C_H3$ region of mouse IgG1
5' GCTACTAGTTTTTGATTTCCACCTTGG 3'	3' primer for amplification of $V_L$
5' CAGCCATGGCCGACATCCAGATG 3'	5' primer for amplification of $V_L$
5' AATTTTACTAGTCACCTTGGTGTGCTGCTGGC 3'	Unique 3' primer for amplification of $V_H$ including part of the mouse gamma 1 first constant
5' TATGCAACTAGTACAACCACAATCCCTGGGCACAATTTT 3'	Unique 3' primer for amplification of $V_H$ including part of mouse gamma 1 first constant region and hinge region

TABLE 2

5' CCAGTTCGAGCTCGTTGTGACTCAGGAATCT 3'	Unique 5' primer for the amplification of $V_L$
5' CCAGTTCGAGCTCGTTGTGACGCGCGCCC 3'	"
5' CCAGTTCGAGCTCGTTGTCTACCCAGTCTCCA 3'	"
5' CCAGTTCGAGCTCCAGATGACCCAGTCTCCA 3'	"
5' CCAGATGTGAGCTCGTTGATGACCCAGACTCCA 3'	"
5' CCAGATGTGAGCTCGTTGATGACCCAGTCTCCA 3'	"
5' CCAGATGTGAGCTCTTTGATGACCCAAACTCAA 3'	"
5' CCAGATGTGAGCTCGTGATAACCCAGGATGAA 3'	"
5' GCAGCATTTCTAGAGTTTCAGCTCCAGCTTGCC 3'	Unique 3' primer for $V_L$ amplification
5' CCGCCGCTCTAGAACACTCATTCCTGTTGAAGCT 3'	Unique 3' primer for $V_L$ amplification including the kappa constant region
5' CCGCCGCTCTAGAACATTCTGCAGGAGACAGACT 3'	Unique 3' primer for $V_L$ amplification including the lambda constant region
5' CCAGTTCGAGCTCGTGATGACACAGTCTCCA 3'	Unique 5' primer for $V_L$ amplification
5' GCGCCGCTCTAGAAATTAACACTCATTCCTGTTGAA 3'	Unique 3' primer for $V_L$ amplification
5' CTATTAACTAGTAAACGGTAACAGTGGTGCCTTGCCCCCA 3'	"

TABLE 2 (Continued)

5) 5' AGGCTTACTAGTACAATCCCCTGGGCACAAT 3'	Unique 3' primer for V <sub>H</sub> amplification
6) 5' GCCGCTCTAGAACACTCATTCCTGTTGAA 3'	Unique 3' primer for V <sub>L</sub> amplification
7) 5' AGGTIIAICTICTCGAGTCTGC 3'	Degenerate 5' primer containing inosine at 4 degenerate positions
8) 5' AGGTIIAICTICTCGAGTCAGC 3'	" "

## 2. Production Of A V<sub>H</sub> Coding Repertoire Enriched In FITC Binding Proteins

Fluorescein isothiocyanate (FITC) was selected as a ligand for receptor binding. It was further decided to enrich by immunization the immunological gene repertoire, i.e., V<sub>H</sub>- and V<sub>L</sub>-coding gene repertoires, for genes coding for anti-FITC receptors. This was accomplished by linking FITC to keyhole limpet hemocyanin (KLH) using the techniques described in Antibodies A Laboratory Manual, Harlow and Lowe, eds., Cold Spring Harbor, New York, (1988). Briefly, 10.0 milligrams (mg) of keyhole limpet hemocyanin and 0.5 mg of FITC were added to 1 ml of buffer containing 0.1 M sodium carbonate at pH 9.6 and stirred for 18 to 24 hours at 4 degrees C (4C). The unbound FITC was removed by gel filtration through Sephadex G-25.

The KLH-FITC conjugate was prepared for injection into mice by adding 100  $\mu$ g of the conjugate to 250  $\mu$ l of phosphate buffered saline (PBS). An equal volume of complete Freund's adjuvant was added and emulsified the entire solution for 5 minutes. A 129 G<sub>1x</sub>+ mouse was injected with 300  $\mu$ l of the emulsion. Injections were given subcutaneously at several sites using a 21 gauge needle. A second immunization with KLH-FITC was given two weeks later. This injection was prepared as follows: fifty  $\mu$ g of KLH-FITC were diluted in 250  $\mu$ L of PBS and an equal volume of alum was admixed to the KLH-FITC solution. The mouse was injected intraperitoneally with 500  $\mu$ l of the solution using a 23 gauge needle. One month later the mice were given a final injection of 50  $\mu$ g of the KLH-FITC conjugate diluted to 200  $\mu$ L in PBS. This injection was given intravenously in the lateral tail vein using a 30 gauge needle. Five days after



This final injection the mice were sacrificed and total cellular RNA was isolated from their spleens.

Hybridoma PCP 8D11 producing an antibody immunospecific for phosphonate ester was cultured in DMEM media (Gibco Laboratories, Grand Island, New York) containing 10 percent fetal calf serum supplemented with penicillin and streptomycin. About  $5 \times 10^8$  hybridoma cells were harvested and washed twice in phosphate buffered saline. Total cellular RNA was prepared from these isolated hybridoma cells.

### 3. Preparation Of A $V_H$ -Coding Gene Repertoire

Total cellular RNA was prepared from the spleen of a single mouse immunized with KLH-FITC as described in Example 2 using the RNA preparation methods described by Chomczynski et al., Anal Biochem., 162:156-159 (1987) using the manufacturer's instructions and the RNA isolation kit produced by Stratagene Cloning Systems, La Jolla, CA. Briefly, immediately after removing the spleen from the immunized mouse, the tissue was homogenized in 10 ml of a denaturing solution containing 4.0 M guanine isothiocyanate, 0.25 M sodium citrate at pH 7.0, and 0.1 M 2-mercaptoethanol using a glass homogenizer. One ml of sodium acetate at a concentration of 2 M at pH 4.0 was admixed with the homogenized spleen. One ml of phenol that had been previously saturated with  $H_2O$  was also admixed to the denaturing solution containing the homogenized spleen. Two ml of a chloroform:isoamyl alcohol (24:1 v/v) mixture was added to this homogenate. The homogenate was mixed vigorously for ten seconds and maintained on ice for 15 minutes. The homogenate was then transferred to a thick-walled 50 ml polypropylene centrifuged tube

(Fisher Scientific Company, Pittsburgh, PA). The solution was centrifuged at 10,000 x g for 20 minutes at 4C. The upper RNA-containing aqueous layer was transferred to a fresh 50 ml polypropylene centrifuge tube and mixed with an equal volume of isopropyl alcohol. This solution was maintained at -20C for at least one hour to precipitate the RNA. The solution containing the precipitated RNA was centrifuged at 10,000 x g for twenty minutes at 4C. The pelleted total cellular RNA was collected and dissolved in 3 ml of the denaturing solution described above. Three ml of isopropyl alcohol was added to the resuspended total cellular RNA and vigorously mixed. This solution was maintained at -20C for at least 1 hour to precipitate the RNA. The solution containing the precipitated RNA was centrifuged at 10,000 x g for ten minutes at 4C. The pelleted RNA was washed once with a solution containing 75% ethanol. The pelleted RNA was dried under vacuum for 15 minutes and then resuspended in dimethyl pyrocarbonate (DEPC) treated (DEPC-H<sub>2</sub>O) H<sub>2</sub>O.

Messenger RNA (mRNA) enriched for sequences containing long poly A tracts was prepared from the total cellular RNA using methods described in Molecular Cloning A Laboratory Manual, Maniatias et al., eds., Cold Spring Harbor Laboratory, New York, (1982). Briefly, one half of the total RNA isolated from a single immunized mouse spleen prepared as described above was resuspended in one ml of DEPC-H<sub>2</sub>O and maintained at 65C for five minutes. One ml of 2x high salt loading buffer consisting of 100 mM Tris-HCL, 1 M sodium chloride, 2.0 mM disodium ethylene diamine tetraacetic acid (EDTA) at pH 7.5, and 0.2% sodium dodecyl sulfate (SDS) was added to the resuspended RNA and the mixture allowed to cool to

room temperature. The mixture was then applied to an oligo-dT (Collaborative Research Type 2 or Type 3) column that was previously prepared by washing the oligo-dT with a solution containing 0.1 M sodium hydroxide and 5 mM EDTA and then equilibrating the column with DEPC-H<sub>2</sub>O. The eluate was collected in a sterile polypropylene tube and reapplied to the same column after heating the eluate for 5 minutes at 65C. The oligo dT column was then washed with 2 ml of high salt loading buffer consisting of 50 mM Tris-HCL at pH 7.5, 500 mM sodium chloride, 1 mM EDTA at pH 7.5 and 0.1% SDS. The oligo dT column was then washed with 2 ml of 1 X medium salt buffer consisting of 50 mM Tris-HCL at pH 7.5, 100 mM sodium chloride 1 mM EDTA and 0.1% SDS. The messenger RNA was eluted from the oligo dT column with 1 ml of buffer consisting of 10 mM Tris-HCL at pH 7.5, 1 mM EDTA at pH 7.5 and 0.05% SDS. The messenger RNA was purified by extracting this solution with phenol/chloroform followed by a single extraction with 100% chloroform. The messenger RNA was concentrated by ethanol precipitation and resuspended in DEPC H<sub>2</sub>O.

The messenger RNA isolated by the above process contains a plurality of different V<sub>H</sub> coding polynucleotides, i.e., greater than about 10<sup>4</sup> different V<sub>H</sub>-coding genes.

#### 4. Preparation Of A Single V<sub>H</sub> Coding Polynucleotide

Polynucleotides coding for a single V<sub>H</sub> were isolated according to Example 3 except total cellular RNA was extracted from monoclonal hybridoma cells prepared in Example 2. The polynucleotides isolated in this manner code for a single V<sub>H</sub>.

### 5. DNA Homolog Preparation

In preparation for PCR amplification, mRNA prepared according to the above examples was used as a template for cDNA synthesis by a primer extension reaction. In a typical 50 ul transcription reaction, 5-10 ug of spleen or hybridoma mRNA in water was first hybridized (annealed) with 500 ng (50.0 pmol) of the 3' V<sub>H</sub> primer (primer 12, Table 1), at 65C for five minutes. Subsequently, the mixture was adjusted to 1.5 mM dATP, dCTP, dGTP and dTTP, 40 mM Tris-HCl at pH 8.0, 8 mM MgCl<sub>2</sub>, 50 mM NaCl, and 2 mM spermidine. Moloney-Murine Leukemia virus Reverse transcriptase (Stratagene Cloning Systems), 26 units, was added and the solution was maintained for 1 hour at 37C.

PCR amplification was performed in a 100 ul reaction containing the products of the reverse transcription reaction (approximately 5 ug of the cDNA/RNA hybrid), 300 ng of 3' V<sub>H</sub> primer (primer 12 of Table 1), 300 ng each of the 5' V<sub>H</sub> primers (primer 2-10 of Table 1) 200 mM of a mixture of dNTP's, 50 mM KCl, 10 mM Tris-HCl pH 8.3, 15 mM MgCl<sub>2</sub>, 0.1% gelatin and 2 units of Taq DNA polymerase. The reaction mixture was overlaid with mineral oil and subjected to 40 cycles of amplification. Each amplification cycle involved denaturation at 92C for 1 minute, annealing at 52C for 2 minutes and polynucleotide synthesis by Primer extension (elongation) at 72C for 1.5 minutes. The amplified V<sub>H</sub>-coding DNA homolog containing samples were extracted twice with phenol/chloroform, once with chloroform, ethanol precipitated and were stored at -70C in 10 mM Tris-HCl, (pH, 7.5) and 1 mM EDTA.

Using unique 5' primers (2-9, Table 1), efficient V<sub>H</sub>-coding DNA homolog synthesis and

amplification from the spleen mRNA was achieved as shown in Figure 3, lanes R17-R24. The amplified cDNA ( $V_H$ -coding DNA homolog) is seen as a major band of the expected size (360 bp). The intensities of the amplified  $V_H$ -coding polynucleotide fragment in each reaction appear to be similar, indicating that all of these primers are about equally efficient in initiating amplification. The yield and quality of the amplification with these primers was reproducible.

The primer containing inosine also synthesized amplified  $V_H$ -coding DNA homologs from spleen mRNA reproducibly, leading to the production of the expected sized fragment, of an intensity similar to that of the other amplified cDNAs (Figure 4, lane R16). This result indicated that the presence of inosine also permits efficient DNA homolog synthesis and amplification. Clearly indicating how useful such primers are in generating a plurality of  $V_H$ -coding DNA homologs. Amplification products obtained from the constant region primers (primers 11 and 13, Table 1) were more intense indicating that amplification was more efficient, possibly because of a higher degree of homology between the template and primers (Figure 4, Lane R9). Based on these results, a  $V_H$ -coding gene library was constructed from the products of eight amplifications, each performed with a different 5' primer. Equal portions of the products from each primer extension reaction were mixed and the mixed product was then used to generate a library of  $V_H$ -coding DNA homolog-containing vectors.

DNA homologs of the  $V_L$  were prepared from the purified mRNA prepared as described above. In preparation for PCR amplification, mRNA prepared

according to the above examples was used as a template for cDNA synthesis. In a typical 50 ul transcription reaction, 5-10 ug of spleen or hybridoma mRNA in water was first annealed with 300 ng (50.0 pmol) of the 3' V<sub>L</sub> primer (primer 14, Table 1), at 65C for five minutes. Subsequently, the mixture was adjusted to 1.5 mM dATP, dCTP, dGTP, and dTTP, 40 mM Tris-HCL at pH 8.0, 8 mM MgCl<sub>2</sub>, 50 mM NaCl, and 2 mM spermidine. Moloney-Murine Leukemia virus reverse transcriptase (Stratagene Cloning Systems), 26 units, was added and the solution was maintained for 1 hour at 37C. The PCR amplification was performed in a 100 ul reaction containing approximately 5 ug of the cDNA/RNA hybrid produced as described above, 300 ng of the 3' V<sub>L</sub> primer (primer 14 of Table 1), 300 ng of the 5' V<sub>L</sub> primer (primer 15 of Table 1), 200 mM of a mixture of dNTP's, 50 mM KCl, 10 mM Tris-HCl pH 8.3, 15 mM MgCl<sub>2</sub>, 0.1% gelatin and 2 units of Taq DNA polymerase. The reaction mixture was overlaid with mineral oil and subjected to 40 cycles of amplification. Each amplification cycle involved denaturation at 92C for 1 minute, annealing at 52C for 2 minutes and elongation at 72C for 1.5 minutes. The amplified samples were extracted twice with phenol/chloroform, once with chloroform, ethanol precipitated and were stored at -70C in 10 mM Tris-HCl at 7.5 and 1 mM EDTA.

#### 6. Inserting DNA Homologs Into Vectors

In preparation for cloning a library enriched in V<sub>H</sub> sequences, PCR amplified products (2.5 mg/30 ul of 150 mM NaCl, 8 mM Tris-HCl (pH 7.5), 6 mM MgSO<sub>4</sub>, 1 mM DTT, 200 mg/ml bovine serum albumin (BSA) at 37C were digested with restriction enzymes Xho I (125 units) and EcoR I (10 U) and purified on a 1%

agarose gel. In cloning experiments which required a mixture of the products of the amplification reactions, equal volumes (50 ul, 1-10 ug concentration) of each reaction mixture were combined after amplification but before restriction digestion. After gel electrophoresis of the digested PCR amplified spleen mRNA, the region of the gel containing DNA fragments of approximately 350 bps was excised, electroeluted into a dialysis membrane, ethanol precipitated and resuspended in 10 mM Tris-HCl pH 7.5 and 1 mM EDTA to a final concentration of 10 ng/ul. Equimolar amounts of the insert were then ligated overnight at 5C to 1 ug of Lambda ZAP<sup>TM</sup> II vector (Stratagene Cloning Systems, La Jolla, CA) previously cut by EcoR I and Xho I. A portion of the ligation mixture (1 ul) was packaged for 2 hours at room temperature using Gigapack Gold packaging extract (Stratagene Cloning Systems, La Jolla, CA), and the packaged material was plated on XL1-blue host cells. The library was determined to consist of  $2 \times 10^7$  V<sub>H</sub> homologs with less than 30% non-recombinant background.

The vector used above, Lambda Zap II is a derivative of the original Lambda Zap (ATCC # 40,298) that maintains all of the characteristics of the original Lambda Zap including 6 unique cloning sites, fusion protein expression, and the ability to rapidly excise the insert in the form of a phagemid (Bluescript SK-), but lacks the SAM 100 mutation, allowing growth on many Non-Sup F strains, including XL1-Blue. The Lambda Zap II was constructed as described in Short et al., Nucleic Acids Res., 16:7583-7600, 1988, by replacing the Lambda S gene contained in a 4254 base pair (bp) DNA fragment produced by digesting Lambda Zap with the restriction

enzyme NcoI. This 4254 bp DNA fragment was replaced with the 4254 bp DNA fragment containing the Lambda S gene isolated from Lambda gt10 (ATCC # 40,179) after digesting the vector with the restriction enzyme NcoI. The 4254 bp DNA fragment isolated from lambda gt10 was ligated into the original Lambda Zap vector using T4 DNA ligase and standard protocols for such procedures described in Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley and Sons, New York, 1987.

In preparation of cloning a library enriched in  $V_L$  sequences, 2 ug of PCR amplified products (2.5 mg/30 ul of 150 mM NaCl, 8 mM Tris-HCL (pH 7.5), 6 mM  $Mg SO_4$ , 1 mM DTT, 200 mg/ml BSA. 37C) were digested with restriction enzymes Nco I (30 units) and Spe I (45 units). The digested PCR amplified products were purified on a 1% agarose gel using standard electroelution technique described in Molecular Cloning A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor, New York, (1982). Briefly, after gel electroelution of the digested PCR amplified product the region of the gel containing the  $V_L$ -coding DNA fragment of the appropriate size was excised, electroelution into a dialysis membrane, ethanol precipitated and resuspended at a final concentration of 10 ng per ml in a solution containing 10 mM Tris-HCL at pH 7.5 and 1 mM EDTA.

An equal molar amount of DNA representing a plurality of different  $V_L$ -coding DNA homologs was ligated to a pBluescript SK- phagemid vector that had been previously cut with Nco I and Spe I. A portion of the ligation mixture was transformed using the manufacturer's instructions into Epicurian Coli XL1-Blue competent cells (Stragagene Cloning Systems, La Jolla, CA). The transformant library was determined



to consist of  $1.2 \times 10^3$  colony forming units/ug of  $V_L$  homologs with less than 3% non-recombinant background.

5                    7.    Sequencing of Plasmids from the  $V_H$ -  
                         Coding cDNA Library

                         To analyze the Lambda Zap II phage clones  
the clones were excised from Lambda Zap into plasmids  
according to the manufacture's instructions  
10                    (Stratagene Cloning System, La Jolla, CA). Briefly,  
phage plaques were cored from the agar plates and  
transferred to sterile microfuge tubes containing 500  
 $\mu$ l a buffer containing 50 mM Tris-HCL at pH 7.5, 100  
mM NaCl, 10 mM  $MgSO_4$ , and 0.01% gelatin and 20 uL of  
15                    chloroform.

                         For excisions, 200 ul of the phage stock,  
200 ul of XL1-Blue cells ( $A_{600} = 1.00$ ) and 1 ul of  
R408 helper phage ( $1 \times 10^{11}$  pfu/ml) were incubated at  
37C for 15 minutes. The excised plasmids were  
20                    infected into XL1-Blue cells and plated onto LB  
plates containing ampicillin. Double stranded DNA  
was prepared from the phagemid containing cells  
according to the methods described by Holmes et al.,  
Anal. Biochem., 114:193, (1981). Clones were first  
25                    screened for DNA inserts by restriction digests with  
either Pvu II or Bgl I and clones containing the  
putative  $V_H$  insert were sequenced using reverse  
transcriptase according to the general method  
described by Sanger et al., Proc. Natl. Acad. Sci.,  
30                    USA, 74:5463-5467, (1977) and the specific  
modifications of this method provided in the  
manufacturer's instructions in the AMV reverse  
transcriptase  $^{35}S$ -dATP sequencing kit from Stratagene  
Cloning Systems, La Jolla, CA.

## 8. Characterization Of The Cloned V<sub>H</sub> Repertoire

The amplified products which had been digested with Xho I and EcoR I and cloned into Lambda ZAP, resulted in a cDNA library with  $9.0 \times 10^5$  pfu's. In order to confirm that the library consisted of a diverse population of V<sub>H</sub>-coding DNA homologs, the N-terminal 120 bases of 18 clones, selected at random from the library, were excised and sequenced (Figure 5). To determine if the clones were of V<sub>H</sub> gene origin, the cloned sequences were compared with known V<sub>H</sub> sequences and V<sub>L</sub> sequences. The clones exhibited from 80 to 90% homology with sequences of known heavy chain origin and little homology with sequences of light chain origin when compared with the sequences available in Sequences of Proteins of Immunological Interest by Kabot et al., 4th ed., U.S. Dept. of Health and Human Sciences, (1987). This demonstrated that the library was enriched for the desired V<sub>H</sub> sequence in preference to other sequences, such as light chain sequences.

The diversity of the population was assessed by classifying the sequenced clones into predefined subgroups (Figure 5). Mouse V<sub>H</sub> sequences are classified into eleven subgroups (Figure 5). Mouse V<sub>H</sub> sequences are classified into eleven subgroups [I (A,B), II (A,B,C), III (A,B,C,D), V (A,B)] based on framework amino acid sequences described in Sequences of Proteins of Immunological Interest by Kabot et al., 4th ed., U.S. Dept. of Health and Human Sciences, (1987); Dildrop, Immunology Today, 5:84, (1984); and Brodeur et al., Eur. J. Immunol., 14; 922, (1984). Classification of the sequenced clones demonstrated that the cDNA library contained V<sub>H</sub> sequences of at least 7 different subgroups. Further, a pairwise

comparison of the homology between the sequenced clones showed that no two sequences were identical at all positions, suggesting that the population is diverse to the extent that it is possible to characterize by sequence analysis.

Six of the clones (L 36-50, Figure 5) belong to the subclass III B and had very similar nucleotide sequences. This may reflect a preponderance of mRNA derived from one or several related variable genes in stimulated spleen, but the data does not permit ruling out the possibility of a bias in the amplification process.

#### 9. V<sub>H</sub>-Expression Vector Construction

The main criterion used in choosing a vector system was the necessity of generating the largest number of Fab fragments which could be screened directly. Bacteriophage lambda was selected as the expression vector for three reasons. First, in vitro packaging of phage DNA is the most efficient method of reintroducing DNA into host cells. Second, it is possible to detect protein expression at the level of single phage plaques. Finally, the screening of phage libraries typically involve less difficulty with nonspecific binding. The alternative, plasmid cloning vectors, are only advantageous in the analysis of clones after they have been identified. This advantage is not lost in the present system because of the use of lambda zap, thereby permitting a plasmid containing the heavy chain, light chain, or Fab expressing inserts to be excised.

To express the plurality of V<sub>H</sub>-coding DNA homologs in an E. coli host cell, a vector was constructed that placed the V<sub>H</sub>-coding DNA homologs in the proper reading frame, provided a ribosome binding

site as described by Shine et al., Nature, 254:34, 1975, provided a leader sequence directing the expressed protein to the periplasmic space, provided a polynucleotide sequence that coded for a known epitope (epitope tag) and also provided a polynucleotide that coded for a spacer protein between the V<sub>H</sub>-coding DNA homolog and the polynucleotide coding for the epitope tag. A synthetic DNA sequence containing all of the above polynucleotides and features was constructed by designing single stranded polynucleotide segments of 20-40 bases that would hybridize to each other and form the double stranded synthetic DNA sequence shown in Figure 6. The individual single-stranded polynucleotides (N<sub>1</sub>-N<sub>12</sub>) are shown in Table 3.

Polynucleotides 2, 3, 9-4', 11, 10-5', 6, 7 and 8 were kinased by adding 1  $\mu$ l of each polynucleotide (0.1 ug/ $\mu$ l) and 20 units of T<sub>4</sub> polynucleotide kinase to a solution containing 70 mM Tris-HCL at pH 7.6, 10 mM MgCl<sub>2</sub>, 5 mM DTT, 10 mM 2ME, 500 micrograms per ml of BSA. The solution was maintained at 37C for 30 minutes and the reaction stopped by maintaining the solution at 65C for 10 minutes. The two end polynucleotides 20 ng of polynucleotides N1 and polynucleotides N12, were added to the above kinasing reaction solution together with 1/10 volume of a solution containing 20.0 mM Tris-HCL at pH 7.4, 2.0 mM MgCl<sub>2</sub> and 50.0 mM NaCl. This solution was heated to 70C for 5 minutes and allowed to cool to room temperature, approximately 25C, over 1.5 hours in a 500 ml beaker of water. During this time period all 10 polynucleotides annealed to form the double stranded synthetic DNA insert shown in Figure 6A. The individual polynucleotides were covalently linked to

each other to stabilize the synthetic DNA insert by adding 40  $\mu$ l of the above reaction to a solution containing 50 mM Tris-HCL at pH 7.5, 7 mM  $MgCl_2$ , 1 mM DTT, 1 mM adenosine triphosphate (ATP) and 10 units of T4 DNA ligase. This solution was maintained at 37C for 30 minutes and then the T4 DNA ligase was inactivated by maintaining the solution at 65C for 10 minutes. The end polynucleotides were kinased by mixing 52  $\mu$ l of the above reaction, 4  $\mu$ l of a solution containing 10 mM ATP and 5 units of T4 polynucleotide kinase. This solution was maintained at 37C for 30 minutes and then the T4 polynucleotide kinase was inactivated by maintaining the solution at 65C for 10 minutes. The completed synthetic DNA insert was ligated directly into a lambda Zap II vector that had been previously digested with the restriction enzymes Not I and Xho I. The ligation mixture was packaged according to the manufacture's instructions using Gigapack II Gold packing extract available from Stratagene Cloning Systems, La Jolla, CA. The packaged ligation mixture was plated on XL1 blue cells (Stratagene Cloning Systems, San Diego, CA). Individual lambda Zap II plaques were cored and the inserts excised according to the in vivo excision protocol provided by the manufacturer, Stratagene Cloning Systems, La Jolla, CA. This in vivo excision protocol moves the cloned insert from the lambda Zap II vector into a plasmid vector to allow easy manipulation and sequencing. The accuracy of the above cloning steps was confirmed by sequencing the insert using the Sanger dideoxide method described in by Sanger et al., Proc. Natl. Acad. Sci USA, 74:5463-5467, (1977) and using the manufacture's instructions in the AMV Reverse Transcriptase  $^{35}S$ -ATP sequencing kit from Stratagene Cloning Systems, La Jolla, CA.

The sequence of the resulting  $V_H$  expression vector is shown in Figure 6A and Figure 7.

Table 3

5	N1)	5' GGCCGCAAATTCTATTTCAAGGAGACAGTCAT 3'
	N2)	5' AATGAAATACCTATTGCCTACGGCAGCCGCTGGATT 3'
	N3)	5' GTTATTACTCGCTGCCCAACCAGCCATGGCCC 3'
	N4)	5' AGGTGAAACTGCTCGAGAATTCTAGACTAGGTAAATAG 3'
10	N5)	5' TCGACTATTAAGTAGTCTAGAATTCTCGAG 3'
	N6)	5' CAGTTTCACCTGGGCCATGGCTGGTTGGG 3'
	N7)	5' CAGCGAGTAATAACAATCCAGCGGCTGCCGTAGGCAATAG 3'
	N8)	5' GTATTTTCATTATGACTGTCTCCTTGAAATAGAATTTGC 3'
	N9-4)	5' AGGTGAAACTGCTCGAGATTTCTAGACTAGTTACCCGTAC 3'
15	N11)	5' GACGTTCCGGACTACGGTTCTTAATAGAATTTCG 3'
	N12)	5' TCGACGAATTCTATTAAGAACCGTAGTC 3'
	N10-5)	5' CGGAACGTCGTACGGGTAAGTAGTCTAGAAATCTCGAG 3'

#### 10. $V_L$ Expression Vector Construction

To express the plurality of  $V_L$  coding polynucleotides in an E. coli host cell, a vector was constructed that placed the  $V_L$  coding polynucleotide in the proper reading frame, provided a ribosome binding site as described by Shine et al., Nature, 254:34, (1975), provided a leader sequence directing the expressed protein to the periplasmic space and also provided a polynucleotide that coded for a spacer protein between the  $V_L$  polynucleotide and the polynucleotide coding for the epitope tag. A synthetic DNA sequence containing all of the above polynucleotides and features was constructed by designing single stranded polynucleotide segments of 20-40 bases that would hybridize to each other and form the double stranded synthetic DNA sequence shown in Figure 6B. The individual single-stranded polynucleotides ( $N_1$ - $N_8$ ) are shown in Table 3.

Polynucleotides N2, N3, N4, N6, N7 and N8 were

kinased by adding 1  $\mu$ l of each polynucleotide and 20 units of T<sub>4</sub> polynucleotide kinase to a solution containing 70 mM Tris-HCL at pH 7.6, 10 mM MgCl<sub>2</sub>, 5 mM DDT, 10 mM 2ME, 500 micrograms per ml of BSA. The solution was maintained at 37C for 30 minutes and the reaction stopped by maintaining the solution at 65C for 10 minutes. The two end polynucleotides 20 ng of polynucleotides N1 and polynucleotides N5 were added to the above kinasing reaction solution together with 1/10 volume of a solution containing 20.0 mM Tris-HCL at pH 7.4, 2.0 mM MgCl<sub>2</sub> and 50.0 mM NaCl. This solution was heated to 70 C for 5 minutes and allowed to cool to room temperature, approximately 25C, over 1.5 hours in a 500 ml beaker of water. During this time period all the polynucleotides annealed to form the double stranded synthetic DNA insert. The individual polynucleotides were covalently linked to each other to stabilize the synthetic DNA insert with adding 40  $\mu$ l of the above reaction to a solution containing 50  $\mu$ l Tris-HCL at pH 7.5, 7 mM MgCl<sub>2</sub>, 1 mM DTT, 1 mM ATP and 10 units of T4 DNA ligase. This solution was maintained at 37C for 30 minutes and then the T4 DNA ligase was inactivated by maintaining the solution at 65C for 10 minutes. The end polynucleotides were kinased by mixing 52  $\mu$ l of the above reaction, 4  $\mu$ l of a solution recontaining 10 mM ATP and 5 units of T4 polynucleotide kinase. This solution was maintained at 37C for 30 minutes and then the T4 polynucleotide kinase was inactivated by maintaining the solution at 65C for 10 minutes. The completed synthetic DNA insert was ligated directly into a lambda Zap II vector that had been previously digested with the restriction enzymes Not I and Xho I. The ligation mixture was packaged according to the manufacture's instructions using Gigapack II Gold packing extract available from Stratagene Cloning Systems, La Jolla, CA. The packaged ligation mixture was plated on XL1-Blue cells (Stratagene Cloning Systems, La Jolla, CA). Individual

lambda Zap II plaques were cored and the inserts excised according to the in vivo excision protocol provided by the manufacturer, Stratagene Cloning Systems, La Jolla, CA and described in Short et al., Nucleic Acids Res., 16:7583-7600, 1988. This in vivo excision protocol moves the cloned insert from the lambda Zap II vector into a phagemid vector to allow easy manipulation and sequencing and also produces the phagemid version of the  $V_L$  expression vectors. The accuracy of the above cloning steps was confirmed by sequencing the insert using the Sanger dideoxide method described by Sanger et al., Proc. Natl. Acad. Sci. USA, 74:5463-5467, (1977) and using the manufacturer's instructions in the AMV reverse transcriptase  $^{35}\text{S}$ -dATP sequencing kit from Stratagene Cloning Systems, La Jolla, CA. The sequence of the resulting  $V_L$  expression vector is shown in Figure 6 and Figure 8.

The  $V_L$  expression vector used to construct the  $V_L$  library was the phagemid produced to allow the DNA of the  $V_L$  expression vector to be determined. The phagemid was produced, as detailed above, by the in vivo excision process from the Lambda Zap  $V_L$  expression vector (Figure 8). The phagemid version of this vector was used because the Nco I restriction enzyme site is unique in this version and thus could be used to operatively link the  $V_L$  DNA homologs into the expression vector.

#### 11. $V_L$ II-Expression Vector Construction

To express the plurality of  $V_L$ -coding DNA homologs in an E. coli host cell, a vector was constructed that placed the  $V_L$ -coding DNA homologs in the proper reading frame, provided a ribosome binding site as described by Shine et al., Nature, 254:34, 1975, provided the Pel B gene leader sequence that has been previously used to successfully secrete Fab fragments in E. coli by Lei et al., J. Bac., 169:4379 (1987) and Better et al., Science,



240:1041 (1988), and also provided a polynucleotide containing a restriction endonuclease site for cloning. A synthetic DNA sequence containing all of the above polynucleotides and features was constructed by designing single stranded polynucleotide segments of 20-60 bases that would hybridize to each other and form the double stranded synthetic DNA sequence shown in Figure 10. The sequence of each individual single-stranded polynucleotides (01-08) within the double stranded synthetic DNA sequence is shown in Table 4.

Polynucleotides 02, 03, 04, 05, 06 and 07 were kinased by adding 1  $\mu$ l (0.1 ug/ $\mu$ l) of each polynucleotide and 20 units of T<sub>4</sub> polynucleotide kinase to a solution containing 70 mM Tris-HCL at pH 7.6, 10 mM magnesium chloride (MgCl), 5 mM dithiothreitol (DTT), 10 mM 2-mercaptoethanol (2ME), 500 micrograms per ml of bovine serum albumin. The solution was maintained at 37C for 30 minutes and the reaction stopped by maintaining the solution at 65C for 10 minutes. The 20 ng each of the two end polynucleotides, 01 and 08, were added to the above kinasing reaction solution together with 1/10 volume of a solution containing 20.0 mM Tris-HCL at pH 7.4, 2.0 mM MgCl and 15.0 mM sodium chloride (NaCl). This solution was heated to 70C for 5 minutes and allowed to cool to room temperature, approximately 25C, over 1.5 hours in a 500 ml beaker of water. During this time period all 8 polynucleotides annealed to form the double stranded synthetic DNA insert shown in Figure 9. The individual polynucleotides were covalently linked to each other to stabilize the synthetic DNA insert by adding 40  $\mu$ l of the above reaction to a solution containing 50 ml Tris-HCL at pH 7.5, 7 ml MgCl, 1 mm DTT, 1 mm ATP and 10 units of T<sub>4</sub> DNA ligase. This solution was maintained at 37C for 30 minutes and then the T<sub>4</sub> DNA ligase was inactivated by maintaining the solution at 65C for 10 minutes. The end polynucleotides were kinased by

mixing 52  $\mu$ l of the above reaction, 4  $\mu$ l of a solution containing 10 mM ATP and 5 units of T4 polynucleotide kinase. This solution was maintained at 37C for 30 minutes and then the T4 polynucleotide kinase was inactivated by maintaining the solution at 65C for 10 minutes. The completed synthetic DNA insert was ligated directly into a lambda Zap II vector that had been previously digested with the restriction enzymes Not I and Xho I. The ligation mixture was packaged according to the manufacture's instructions using Gigapack II Gold packing extract available from Stratagene Cloning Systems, La Jolla, CA. The packaged ligation mixture was plated on XL1 blue cells (Stratagene Cloning Systems, San Diego, CA). Individual lambda Zap II plaques were cored and the inserts excised according to the in vivo excision protocol provided by the manufacturer, Stratagene Cloning Systems, La Jolla, CA. This in vivo excision protocol moves the cloned insert from the lambda Zap II vector into a plasmid vector to allow easy manipulation and sequencing. The accuracy of the above cloning steps was confirmed by sequencing the insert using the manufacture's instructions in the AMV Reverse Transcriptase <sup>35</sup>S-dATP sequencing kit from Stratagene Cloning Systems, La Jolla, CA. The sequence of the resulting V<sub>L</sub>II-expression vector is shown in Figure 9 and Figure 11.

TABLE 4

01)	5' TGAATTCTAAACTAGTCGCCAAGGAGACAGTCAT 3'
02)	5' AATGAAATACCTATTGCCTACGGCAGCCGCTGGATT 3'
03)	5' GTTATTACTCGCTGCCCAACCAGCCATGGCC 3'
5 04)	5' GAGCTCGTCAGTTCTAGAGTTAAGCGGCCG 3'
05)	5' GTATTTTATTATGACTGTCTCCTTGGCGACTAGTTTAGAA- TTCAAGCT 3'
06)	5' CAGCGAGTAATAACAATCCAGCGGCTGCCGTAGGCAATAG 3'
07)	5' TGACGAGCTCGGCCATGGCTGGTTGGG 3'
10 08)	5' TCGACGGCCGCTTAACTCTAGAAC 3'

12. V<sub>H</sub> + V<sub>L</sub> Library Construction

To prepare an expression library enriched in V<sub>H</sub> sequences, DNA homologs enriched in V<sub>H</sub> sequences were prepared according to Example 6 using the same set of 5' primers but with primer 12A (Table 1) as the 3' primer. These homologs were then digested with the restriction enzymes Xho I and Spe I and purified on a 1% agarose gel using the standard electroelution technique described in Molecular Cloning A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor, New York, (1982). These prepared V<sub>H</sub> DNA homologs were then directly inserted into the V<sub>H</sub> expression vector that had been previously digested with Xho I and Spe I.

The ligation mixture containing the V<sub>H</sub> DNA homologs were packaged according to the manufacturers specifications using Gigapack Gold II Packing Extract (Stratagene Cloning Systems, La Jolla, CA). The expression libraries were then ready to be plated on XL-1 Blue cells.

To prepare a library enriched in V<sub>L</sub> sequences, PCR amplified products enriched in V<sub>L</sub> sequences were prepared according to Example 6. These V<sub>L</sub> DNA homologs were digested with restriction enzymes Nco I and Spe I. The digested V<sub>L</sub> DNA homologs were purified on a 1% agarose gel using

standard electroelution techniques described in Molecular Cloning A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor, NY (1982). The prepared  $V_L$  DNA homologs were directly inserted into the  $V_L$  expression vector that had been previously digested with the restriction enzymes Nco I and Spe I. The ligation mixture containing the  $V_L$  DNA homologs were transformed into XL-1 blue competent cells using the manufacturer's instructions (Stratagene Cloning Systems, La Jolla, CA).

### 13. Inserting $V_L$ Coding DNA Homologs Into $V_L$ Expression Vector

In preparation for cloning a library enriched in  $V_L$  sequences, PCR amplified products (2.5 ug/30 ul of 150 mM NaCl, 8 mM Tris-HCl (pH 7.5), 6 mM  $MgSO_4$ , 1 mM DTT, 200 ug/ml BSA at 37C were digested with restriction enzymes Sac I (125 units) and Xba I (125 units) and purified on a 1% agarose gel. In cloning experiments which required a mixture of the products of the amplification reactions, equal volumes (50 ul, 1-10 ug concentration) of each reaction mixture were combined after amplification but before restriction digestion. After gel electrophoresis of the digested PCR amplified spleen mRNA, the region of the gel containing DNA fragments of approximate 350 bps was excised, electroeluted into a dialysis membrane, ethanol precipitated and resuspended in a TE solution containing 10 mM Tris-HCl pH 7.5 and 1 mM EDTA to a final concentration of 50 ng/ul.

The  $V_{LII}$ -expression DNA vector was prepared for cloning by admixing 100 ug of this DNA to a solution containing 250 units each of the restriction endonucleases Sac 1 and Xba 1 (both from Boehringer Mannheim, Indianapolis, IN) and a buffer recommended by the manufacturer. This solution was maintained at 37 from 1.5 hours. The solution was heated at 65C for 15 minutes top

inactivate the restriction endonucleases. The solution was chilled to 30C and 25 units of heat-killable (HK) phosphatase (Epicenter, Madison, WI) and  $\text{CaCl}_2$  were admixed to it according to the manufacturer's specifications. This solution was maintained at 30C for 1 hour. The DNA was purified by extracting the solution with a mixture of phenol and chloroform followed by ethanol precipitation. The  $V_L$ II expression vector was now ready for ligation to the  $V_L$  DNA homologs prepared in the above examples.

DNA homologs enriched in  $V_L$  sequences were prepared according to Example 5 but using a 5' light chain primer and the 3' light chain primer shown in Table 2. Individual amplification reactions were carried out using each 5' light chain primer in combination with the 3' light chain primer. These separate  $V_L$  homolog containing reaction mixtures were mixed and digested with the restriction endonucleases Sac 1 and Xba 1 according to Example 6. The  $V_L$  homologs were purified on a 1% agarose gel using the standard electroelution technique described in Molecular Cloning A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor, New York, (1982). These prepared  $V_L$  DNA homologs were then directly inserted into the Sac 1 - Xba cleaved  $V_L$ II-expression vector that was prepared above by ligating 3 moles of  $V_L$  DNA homolog inserts with each mole of the  $V_L$ II-expression vector overnight at 5C.  $3.0 \times 10^5$  plaque forming units were obtained after packaging the DNA with Gigapack II Bold (Stratagene Cloning Systems, La Jolla, CA) and 50% were recombinants.

#### 14. Randomly Combining $V_H$ and $V_L$ DNA Homologs on the Same Expression Vector

The  $V_L$ II-expression library prepared in Example 13 was amplified and 500 ug of  $V_L$ II-expression library phage DNA prepared from the amplified phage stock using the

procedures described in Molecular Cloning: A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982), 50 ug of this  $V_{LII}$ -expression library phage DNA was maintained in a solution containing 100 units of MluI restriction endonuclease (Boehringer Mannheim, Indianapolis, IN) in 200 ul of a buffer supplied by the endonuclease manufacturer for 1.5 hours at 37C. The solution was then extracted with a mixture of phenol and chloroform. The DNA was then ethanol precipitated and resuspended in 100 ul of water. This solution was admixed with 100 units of the restriction endonuclease EcoR I (Boehringer Mannheim, Indianapolis, IN) in a final volume of 200 ul of buffer containing the components specified by the manufacturer. This solution was maintained at 37C for 1.5 hours and the solution was then extracted with a mixture of phenol and chloroform. The DNA was ethanol precipitated and the DNA resuspended in TE.

The  $V_H$  expression library prepared in Example 12 was amplified and 500 ug of  $V_H$  expression library phage DNA prepared using the methods detailed above. 50 ug of the  $V_H$  expression library phage DNA was maintained in a solution containing 100 units of Hind III restriction endonuclease (Boehringer Mannheim, Indianapolis, IN) in 200 ul of a buffer supplied by the endonuclease manufacturer for 1.5 hours at 37C. The solution was then extracted with a mixture of phenol and chloroform saturated with 0.1 M Tris-HCL at pH 7.5. The DNA was then ethanol precipitated and resuspended in 100 ul of water. This solution was admixed with 100 units of the restriction endonuclease EcoR I (Boehringer Mannheim, Indianapolis, IN) in a final volume of 200 ul of buffer containing the components specified by the manufacturer. This solution was maintained at 37C for 1.5 hours and the solution was then extracted with a mixture of phenol and chloroform. The DNA was ethanol precipitated and the DNA resuspended in TE.

The restriction digested  $V_H$  and  $V_L$ II-expression Libraries were ligated together. The ligation reaction consisted of 1 ug of  $V_H$  and 1 ug of  $V_L$ II phage library DNA in a 10 ul reaction using the reagents supplied in a ligation kit purchased from Stratagene Cloning Systems (La Jolla, California). After ligation for 16 hr at 4C, 1 ul of the ligated the phage DNA was packaged with Gigapack Gold II packaging extract and plated on XL 1-blue cells prepared according to the manufacturers instructions. A portion of the  $3 \times 10^6$  clones obtained were used to determine the effectiveness of the combination. The resulting  $V_H$  and  $V_L$  expression vector is shown in Figure 11.

Clones containing both  $V_H$  and  $V_L$  were excised from the phage to pBluescript using the in vitro excision protocol described by Short et al., Nucleic Acid Research, 16:7583-7600 (1988). Clones chosen for excision expressed the decapetide tag and did not cleave X-gal in the presence of 2mM IPTG thus remaining white. Clones with these characteristics represented 30% of the library. 50% of the clones chosen for excision contained a  $v_H$  and  $V_L$  as determined by restriction analysis. Since approximately 30% of the clones in the  $V_H$  library expressed the decapetide tag and 50% of the clones in the  $V_L$ II library contained a  $V_L$  sequence it was anticipated that no more than 15% of the clones in the combined library would contain both  $V_H$  and  $V_L$  clones. The actual number obtained was 15% of the library indicating that the process of combination was very efficient.

#### 15. Segregating DNA Homologs For a $V_H$ Antigen Binding Protein

To segregate the individual clones containing DNA homologs that code for a  $V_H$  antigen binding protein, the titre of the  $V_H$  expression library prepared according to Example 11 was determined. This library titration was

performed using methods well known to one skilled in the art. Briefly, serial dilutions of the library were made into a buffer containing 100 mM NaCl, 50 mM Tris-HCL at pH 7.5 and 10 mM MgSO<sub>4</sub>. Ten ul of each dilution was added to 200 ul of exponentially growing E. coli cells and maintained at 37C for 15 minutes to allow the phage to absorb to the bacterial cells. Three ml of top agar consisting of 5 g/L NaCl, 2 g/L of MgSO<sub>4</sub>, 5 g/L yeast extract, 10 g/L NZ amine (casein hydrolysate) and 0.7% melted, 50C agarose. The phage, the bacteria and the top agar were mixed and then evenly distributed across the surface of a prewarmed bacterial agar plate (5 g/L NaCl, 2 g/L MgSO<sub>4</sub>, 5 g/L yeast extract, 10 g/L NZ amine (casein hydrolysate) and 15 g/L Difco agar. The plates were maintained at 37C for 12 to 24 hours during which time period the lambda plaques developed on the bacterial lawn. The lambda plaques were counted to determine the total number of plaque forming units per ml in the original library.

The titred expression library was then plated out so that replica filters could be made from the library. The replica filters will be used to later segregate out the individual clones in the library that are expressing the antigens binding proteins of interest. Briefly, a volume of the titred library that would yield 20,000 plaques per 150 millimeter plate was added to 600 ul of exponentially growing E. coli cells and maintained at 37C for 15 minutes to allow the phage to absorb to the bacterial cells. Then 7.5 ml of top agar was admixed to the solution containing the bacterial cells and the absorbed phage and the entire mixture distributed evenly across the surface of a prewarmed bacterial agar plate. This process was repeated for a sufficient number of plates to plate out a total number of plaques at least equal to the library size. These plates were then maintained at 37 C for 5 hours. The plates were then overlaid with nitrocellulose filters that had been



pretreated with a solution containing 10 mM isopropyl-beta-D-thiogalactopyranosid (IPTG) and maintained at 37C for 4 hours. The orientation of the nitrocellulose filters in relation to the plate were marked by punching a hole with a needle dipped in waterproof ink through the filter and into the bacterial plates at several locations. The nitrocellulose filters were removed with forceps and washed once in a TBST solution containing 20 mM Tris-HCL at pH 7.5, 150 mM NaCl and 0.05% monolaurate (tween-20). A second nitrocellulose filter that had also been soaked in a solution containing 10 mM IPTG was reapplied to the bacterial plates to produce duplicate filters. The filters were further washed in a fresh solution of TBST for 15 minutes. Filters were then placed in a blocking solution consisting of 20 mM Tris-HCL at pH 7.5, 150 mM NaCl and 1% BSA and agitated for 1 hour at room temperature. The nitrocellulose filters were transferred to a fresh blocking solution containing a 1 to 500 dilution of the primary antibody and gently agitated for at least 1 hour at room temperature. After the filters were agitated in the solution containing the primary antibody the filters were washed 3 to 5 times in TBST for 5 minutes each time to remove any of the residual unbound primary antibody. The filters were transferred into a solution containing fresh blocking solution and a 1 to 500 to a 1 to 1,000 dilution of alkaline phosphatase conjugated secondary antibody. The filters were gently agitated in the solution for at least 1 hour at room temperature. The filters were washed 3 to 5 times in a solution of TBST for at least 5 minutes each time to remove any residual unbound secondary antibody. The filters were washed once in a solution containing 20 mM Tris-HCL at pH 7.5 and 150 mM NaCl. The filters were removed from this solution and the excess moisture blotted from them with filter paper. The color was developed by placing the filter in a solution containing 100 mM Tris-HCL

at pH 9.5, 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 0.3 mg/ml of nitro Blue Tetrazolium (NBT) and 0.15 mg/ml of 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) for at least 30 minutes at room temperature. The residual color development solution was  
5 rinsed from the filter with a solution containing 20 mM Tris-HCL at pH 7.5 and 150 mM NaCl. The filter was then placed in a stop solution consisting of 20 mM Tris-HCL at pH 2.9 and 1 mM EDTA. The development of an intense purple color indicates at positive result. The filters are used to  
10 locate the phage plaque that produced the desired protein. That phage plaque is segregated and then grown up for further analysis.

Several different combinations of primary antibodies and second antibodies were used. The first  
15 combination used a primary antibody immunospecific for a decapeptide that will be expressed only if the V<sub>H</sub> antigen binding protein is expressed in the proper reading frame to allow read through translation to include the decapeptide epitope covalently attached to the V<sub>H</sub> antigen binding  
20 protein. This decapeptide epitope and an antibody immunospecific for this decapeptide epitope was described by Green et al., Cell 28:477 (1982) and Niman et al., Proc. Nat. Acad. Sci. U.S.A. 80:4949 (1983). The sequence of the decapeptide recognized is shown in Figure 11. A functional  
25 equivalent of the monoclonal antibody that is immunospecific for the decapeptide can be prepared according to the methods of Green et al. and Niman et al. The secondary antibody used with this primary antibody was a goat antimouse IgG (Fisher Scientific). This antibody was immunospecific for  
30 the constant region of mouse IgG and did not recognize any portion of the variable region of heavy chain. This particular combination of primary and secondary antibodies when used according to the above protocol determined that between 25% and 30% of the clones were expressing the  
35 decapeptide and therefore these clones were assumed to also

be expressing a  $V_H$  antigen binding protein.

In another combination the anti-decapeptide mouse monoclonal was used as the primary antibody and an affinity purified goat anti-mouse Ig, commercially available as part of the picoBlue immunoscreening kit from Stratagene Cloning System, La Jolla, CA, was use as the secondary antibody. This combination resulted in a large number of false positive clones because the secondary antibody also immunoreacted with the  $V_H$  of the heavy chain. Therefore this antibody reacted with all clones expressing any  $V_H$  protein and this combination of primary and secondary antibodies did not specifically detect clones with the  $V_H$  polynucleotide in the proper reading frame and thus allowing expressing of the decapeptide.

Several combinations of primary and secondary antibodies are used where the primary antibody is conjugated to fluorescein isothiocyanate (FITC) and thus the immunospecificity of the antibody was not important because the antibody is conjugated to the preselected antigen (FITC) and it is that antigen that should be bound by the  $V_H$  antigen binding proteins produced by the clones in the expression library. After this primary antibody has bound by virtue that is FITC conjugated mouse monoclonal antibody p2 5764 (ATCC #HB-9505). The secondary antibody used with this primary antibody is a goat anti-mouse Ig<sup>6</sup> (Fisher Scientific, Pittsburg, PA) conjugated to alkaline phosphatase. Using the method described in Antibodies A Laboratory Manual, Harlow and Lowe, eds., Cold Spring Harbor, New York, (1988). If a particular clone in the  $V_H$  expression, library, expresses a  $V_H$  binding protein that binds the FITC covalently coupled to the primary antibody, the secondary antibody binds specifically and when developed the alkaline phosphate causes a distinct purple color to form.

The second combination of antibodies of the type

uses a primary antibody that is FITC conjugated rabbit anti-human IgG (Fisher Scientific, Pittsburgh, PA). The secondary antibody used with this primary antibody is a goat anti-rabbit IgG conjugated to alkaline phosphatase using the methods described in Antibodies A Laboratory Manual, Harlow and Lane, eds., Cold Spring Harbor, New York, (1988). If a particular clone in the  $V_H$  expression library expresses a  $V_H$  binding protein that binds the FITC conjugated to the primary antibody, the secondary antibody binds specifically and when developed the alkaline phosphatase causes a distinct purple color to form.

Another primary antibody was the mouse monoclonal antibody p2 5764 (ATCC # HB-9505) conjugated to both FITC and  $^{125}\text{I}$ . The antibody would be bound by any  $V_H$  antigen binding proteins expressed. Then because the antibody is also labeled with  $^{125}\text{I}$ , an autoradiogram of the filter is made instead of using a secondary antibody that is conjugated to alkaline phosphatase. This direct production of an autoradiogram allows segregation of the clones in the library expressing a  $V_H$  antigen binding protein of interest.

16. Segregating DNA Homologs For a  
 $V_H$  and  $V_L$  that Form an Antigen  
Binding  $F_V$

To segregate the individual clones containing DNA homologs that code for a  $V_H$  and a  $V_L$  that form an antigen binding  $F_V$  the  $V_H$  and  $V_L$  expression library was titred according to Example 15. The titred expression library was then screened for the presence of the decapetide tag expressed with the  $V_H$  using the methods described in Example 15. DNA was then prepared from the clones to express the decapetide tag. This DNA was digested with the restriction endonuclease Pvu II to determine whether these clones also contained a  $V_L$  DNA homolog. The slower migration of a PvuII restriction endonuclease fragment indicated that the

particular clone contained both a  $V_H$  and a  $V_L$  DNA homolog.

The clones containing both a  $V_H$  and a  $V_L$  DNA homolog were analyzed to determine whether these clones produced an assembled  $F_V$  protein molecule from the  $V_H$  and  $V_L$  DNA homologs.

The  $F_V$  protein fragment produced in clones containing both  $V_H$  and  $V_L$  was visualized by immune precipitation of radiolabeled protein expressed in the clones. A 50 ml culture of LB broth (5 g/L yeast extract, 10 g/L and tryptone 10 g/L NaCl at pH 7.0) containing 100 ug/ul of ampicillin was inoculated with E. Coli harboring a plasmid contain a  $V_H$  and a  $V_L$ . The culture was maintained at 37C with shaking until the optical density measured at 550 nm was 0.5 culture then was centrifuged at 3,000 g for 10 minutes and resuspended in 50 ml of M9 media (6 g/L  $Na_2HPO_4$ , 3 g/L  $KH_2PO_4$ , 0.5 g/L NaCl, 1 g/L  $NH_4Cl$ , 2g/L glucose, 2 mM  $MgSO_4$  and 0.1 mM  $MgSO_4$   $CaCl_2$  supplemented with amino acids without methionine or cysteine. This solution was maintained at 37C for 5 minutes and then 0.5 mCi of  $^{35}S$  as  $HSO_4^-$  (New England Nuclear, Boston, MA) was added and the solution was further maintained at 37C for an additional 2 hours. The solution was then centrifuged at 3000xg and the supernatant discarded. The resulting bacterial cell pellet was frozen and thawed and then resuspended in a solution containing 40 mM Tris pH 8.0, 100 mM sucrose and 1 mM EDTA. The solution was centrifuged at 10000xg for 10 minutes and the resulting pellet discarded. The supernatant was admixed with 10 ul of anti-decapeptide monoclonal antibody and maintained for 30-90 minutes on ice. 40 ul of protein G coupled to sepharose beads (Pharmacia, Piscataway, NJ) was admixed to the solution and the added solution maintained for 30 minutes on ice to allow an immune precipitate to form. The solution was centrifuged at 10,000 xg for 10 minutes and the resulting pellet was resuspended in 1 ml of a solution containing 100 mM Tris-HCL at pH 7.5 and

centrifuged at 10,000 xg for 10 minutes. This procedure was repeated twice. The resulting immune precipitate pellet was loaded onto a PhastGel Homogenous 20 gel (Pharmacia, Piscataway, NJ) according to the manufacturer's directions. The gel was dried and used to expose X-ray film.

The resulting autoradiogram is shown in Figure 12. The presence of assembled  $F_V$  molecules can be seen by the presence of  $V_L$  that was immunoprecipitated because it was attached to the  $V_H$ -decapeptide tag recognized by the precipitating antibody.

#### 17. Generation of a Large Combinatorial Library of the Immunoglobulin Repertoire in Phage

Vectors suitable for expression of  $V_H$ ,  $V_L$ ,  $F_V$  and Fab sequences are diagrammed in Figures 7 and 9. As previously discussed, the vectors were constructed by modification of Lambda Zap by inserting synthetic oligonucleotides into the multiple cloning site. The vectors were designed to be antisymmetric with respect to the Not I and EcoR I restriction sites which flank the cloning and expression sequences. As described below, this antisymmetry in the placement of restriction sites in a linear vector like bacteriophage is the essential feature of the system which allows a library expressing light chains to be combined with one expressing heavy chains to construct combinatorial Fab expression libraries. Lambda Zap II  $V_L$ II (Figure 9) is designed to serve as a cloning vector for light chain fragments and Lambda Zap II  $V_H$  (Figure 7) is designed to serve as a cloning vector for heavy chain sequences in the initial step of library construction. These vectors are engineered to efficiently clone the products of PCR amplification with specific restriction sites incorporated at each end.

#### A. PCR Amplification of Antibody Fragments

The PCR amplification of mRNA isolated from spleen cells with oligonucleotides which incorporate restriction sites into the ends of the amplified product can be used to clone and express heavy chain sequences including Fd and kappa chain sequences. The oligonucleotide primers used for these amplifications are presented in Tables 1 and 2. The primers are analogous to those which have been successfully used in Example 5 for amplification of  $V_H$  sequences. The set of 5' primers for heavy chain amplification were identical to those previously used to amplify  $V_H$  and those for light chain amplification were chosen on similar principles, Sastry et al., Proc. Natl. Acad. Sci. USA, 8G: 5728 (1989) and Orland et al., Proc. Natl. Acad. Sci. USA, 8G:3833 (1989). The unique 3' primers of heavy (IgG1) and light (k) chain sequences were chosen to include the cysteines involved in heavy-light chain disulfide bond formation. At this stage no primer was constructed to amplify lambda light chains since they constitute only a small fraction of murine antibodies. In addition, Fv fragments have been constructed using a 3' primer which is complementary to the mRNA in the J (joining) region (amino acid 128) and a set of unique 5' primers which are complementary to the first strand cDNA in the conserved N-terminal region of the processed protein. Restriction endonuclease recognition sequences are incorporated into the primers to allow for the cloning of the amplified fragment into a lambda phage vector in a predetermined reading frame for expression.

#### B. Library Construction

The construction of a combinatorial library was accomplished in two steps. In the first step, separate heavy and light chain libraries were constructed in Lambda Zap II  $V_H$  and Lambda Zap II  $V_L$  II respectively. In the second step, these two libraries were combined at the

antisymmetric EcoRI sites present in each vector. This resulted in a library of clones each of which potentially co-expresses a heavy and a light chain. The actual combinations are random and do not necessarily reflect the combinations present in the B-cell population in the parent animal. Lambda Zap II V<sub>H</sub> expression vector has been used to create a library of heavy chain sequences from DNA obtained by PCR amplification of mRNA isolated from the spleen of a 129 G<sub>ix</sub> + mouse previously immunized with p-nitrophenyl phosphoramidate (NPN) antigen 1 according to formula I (Figure 13) conjugated to keyhole limpet hemocyanin (KLH).

The NPN-KLH conjugate was prepared by admixture of 250 ul of a solution containing 2.5 mg of NPN according to formula 1 (Figure 13) in dimethylformamide with 750 ul of a solution containing 2 mg of KLH in 0.01 M sodium phosphate buffer (pH 7.2). The two solutions were admixed by slow addition of the NPN solution to the KLH solution while the KLH solution was being agitated by a rotating stirring bar. Thereafter the admixture was maintained at 4° for 1 hour with the same agitation to allow conjugation to proceed. The conjugated NPN-KLH was isolated from the nonconjugated NPN and KLH by gel filtration through Sephadex G-25. The isolated NPN-KLH conjugate was used in mouse immunizations as described in Example 2.

The spleen mRNA resulting from the above immunizations was isolated and used to create a primary library of V<sub>H</sub> gene sequences using the Lambda Zap II V<sub>H</sub> expression vector. The primary library contains  $1.3 \times 10^6$  pfu and has been screened for the expression of the decapeptide tag to determine the percentage of clones expressing Fd sequences. The sequence for this peptide is only in frame for expression following the cloning of a Fd (or V<sub>H</sub>) fragment into the vector. At least 80% of the clones in the library express Fd fragments based on immunodetection of the decapeptide tag.



The light chain library was constructed in the same way as the heavy chain and shown to contain  $2.5 \times 10^6$  members. Plaque screening, using an anti-kappa chain antibody, indicated that 60% of the library contained expressed light chain inserts. This relatively small percentage of inserts probably resulted from incomplete dephosphorylation of vector after cleavage with Sac I and Xba I.

Once obtained, the two libraries were used to construct a combinatorial library by crossing them at the EcoR I site. To accomplish the cross, DNA was first purified from each library. The light chain library was cleaved with MluI restriction endonuclease, the resulting 5' ends dephosphorylated and the product digested with EcoR I. This process cleaved the left arm of the vector into several pieces but the right arm containing the light chain sequences, remained intact. In a parallel fashion, the DNA of heavy chain library was cleaved with HindIII, dephosphorylated and cleaved with EcoR I, destroying the right arm but leaving the left arm containing the heavy chain sequences intact. The DNA's so prepared were then combined and ligated. After ligation only clones which resulted from combination of a right arm of light chain-containing clones and a left arm of heavy chain-containing clones reconstituted a viable phage. After ligation and packaging,  $2.5 \times 10^7$  clones were obtained. This is the combinatorial Fab expression library that was screened to identify clones having affinity for NPN. To determine the frequency the phage clones which co-express the light and heavy chain fragments, duplicate lifts of the light chain, heavy chain and combinatorial libraries were screened as above for light and heavy chain expression. In this study of approximately 500 recombinant phage approximately 60% co-expressed light and heavy chain proteins.

### C. Antigen Binding

All three libraries, the light chain, the heavy chain and Fab were screened to determine if they contained recombinant phage that expressed antibody fragments binding NPN. In a typical procedure 30,000 phage were plated and duplicate lifts with nitrocellulose screened for binding to NPN coupled to  $^{125}\text{I}$  labeled BSA (Figure 15). Duplicate screens of 80,000 recombinant phage from the light chain library and a similar number from the heavy chain library did not identify any clones which bound the antigen. In contrast, the screen of a similar number of clones from the Fab expression library identified many phage plaques that bound NPN (Figure 15). This observation indicates that under conditions where many heavy chains in combination with light chains bind to antigen the same heavy or light chains alone do not. Therefore, in the case of NPN, it is believed that there are many heavy and light chains that only bind antigen when they are combined with specific light and heavy chains respectively.

To assess the ability to screen large numbers of clones and obtain a more quantitative estimate of the frequency of antigen binding clones in the combinatorial library, one million phage plaques were screened and approximately 100 clones which bound to antigen were identified. For six clones which were believed to bind NPN, a region of the plate containing the positive and approximately 20 surrounding bacteriophage plaques was "cored", replated, and screened with duplicate lifts (Figure 15). As expected, approximately one in twenty of the phage specifically bind to antigen. "Cores" of regions of the plated phage believed to be negative did not give positives on replating.

To determine the specificity of the antigen-antibody interaction, antigen binding was competed with free unlabeled antigen as shown in Figure 16. Competition

studies showed that individual clones could be distinguished on the basis of antigen affinity. The concentration of free haptens required for complete inhibition of binding varied between  $10-100 \times 10^9$  M suggesting that the expressed Fab fragments had binding constants in the nanomolar range.

D. Composition of the Clones and Their Expressed Products

In preparation for characterization of the protein products able to bind NPN as described in Example 18C, a plasmid containing the heavy and light chain genes was excised from the appropriate "cored" bacteriophage plaque using M13mp8 helper phage. Mapping of the excised plasmid demonstrated a restriction pattern consistent with incorporation of heavy and light chain sequences. The protein products of one of the clones was analyzed by ELISA and Western blotting to establish the composition of the NPN binding protein. A bacterial supernate following IPTG induction was concentrated and subjected to gel filtration. Fractions in the molecular weight range 40-60 kD were pooled, concentrated and subjected to a further gel filtration separation. As illustrated in Figure 17, ELISA analysis of the eluting fractions demonstrated that NPN binding was associated with a protein of molecular weight about 50 kD which immunological detection showed contained both heavy and light chains. A Western blot (not shown) of a concentrated bacterial supernate preparation under non-reducing conditions was developed with anti-decapeptide antibody. This revealed a protein band of molecular weight of 50 kD. Taken together these results are consistent with NPN binding being a function of Fab fragments in which heavy and light chains are covalently linked.

E. Comparison of the Properties of the In Vivo Repertoire Versus the Phage

Combinatorial Library

In this example a relatively restricted library was prepared because only a limited number of primers were used for PCR amplification of Fd sequences. The library is expected to contain only clones expressing kappa/gamma sequences. However, this is not an inherent limitation of the method since additional primers can be added to amplify any antibody class or subclass. Despite this restriction we were able to isolate a large number of antigen binding clones.

A central issue arising from this work is how a phage library prepared as described herein compares with the in vivo antibody repertoire in terms of size, characteristics of diversity, and ease of access.

The size of the mammalian antibody repertoire is difficult to judge but a figure of the order of  $10^6$ - $10^8$  different antigen specificities is often quoted. With some of the reservations discussed below, a phage library of this size or larger can readily be constructed by a modification of the current method. In fact once an initial combinatorial library has been constructed, heavy and light chains can be shuffled to obtain libraries of exceptionally large numbers.

In principle, the diversity characteristics of the naive (unimmunized) in vivo repertoire and corresponding phage library are expected to be similar in that both involve a random combination of heavy and light chains. However, different factors will act to restrict the diversity expressed by an in vivo repertoire and phage library. For example a physiological modification such as tolerance will restrict the expression of certain antigenic specificities from the in vivo repertoire but these specificities may still appear in the phage library. On the other hand, bias in the cloning process may introduce restrictions into the diversity of the phage library. For example the representation of mRNA for sequences expressed

by stimulated B-cells can be expected to predominate over those of unstimulated cells because of higher levels of expression. Different source tissues (e.g., peripheral blood, bone marrow or regional lymph nodes) and different PCR primers (e.g., ones expected to amplify different antibody classes) may result in libraries with different diversity characteristics.

Another difference between in vivo repertoire and phage library is that antibodies isolated from the former may have benefited from affinity maturation due to somatic mutations after combination of heavy and light chains whereas the latter randomly combines the matured heavy and light chains. Given a large enough phage library derived from a particular in vivo repertoire, the original matured heavy and light chains will be recombined. However, since one of the potential benefits of this new technology is to obviate the need for immunization by the generation of a single highly diverse "generic" phage library, it would be useful to have methods to optimize sequences to compensate for the absence of somatic mutation and clonal selection. Three procedures are made readily available through the methods of the present invention. First, saturation mutagenesis may be performed on the CDR's and the resulting Fabs can be assayed for increased function. Second, a heavy or a light chain of a clone which binds antigen can be recombined with the entire light or heavy chain libraries respectively in a procedure identical to the one used to construct the combinatorial library. Third, iterative cycles of the two above procedures can be performed to further optimize the affinity or catalytic properties of the immunoglobulin. It should be noted that the latter two procedures are not permitted in B-cell clonal selection which suggests that the methods described here may actually increase the ability to identify optimal sequences.

Access is the third area where it is of interest

to compare the in vivo antibody repertoire and phage library. In practical terms the phage library is much easier to access. The screening methods allow one to survey at least 50,000 clones per plate so that  $10^6$  antibodies can be readily examined in a day. This factor alone should encourage the replacement of hybridoma technology with the methods described here. The most powerful screening methods utilize selection which may be accomplished by incorporating selectable markers into the antigen such as leaving groups necessary for replication of auxotrophic bacterial strains or toxic substituents susceptible to catalytic inactivation. There are also further advantages related to the fact that the in vivo antibody repertoire can only be accessed via immunization which is a selection on the basis of binding affinity. The phage library is not similarly restricted. For example, the only general method to identify antibodies with catalytic properties has been by pre-selection on the basis of affinity of the antibody to a transition state analogue. No such restrictions apply to the in vitro library where catalysis can, in principle, be assayed directly. The ability to directly assay large numbers of antibodies for function may allow selection for catalysts in reactions where a mechanism is not well defined or synthesis of the transition state analog is difficult. Assaying for catalysis directly eliminates the bias of the screening procedure for reaction mechanisms pejorative to a synthetic analog and therefore simultaneous exploration of multiple reaction pathways for a given chemical transformation are possible.

The methods disclosed herein describe generation of Fab fragments which are clearly different in a number of important respects from intact (whole) antibodies. There is undoubtedly a loss of affinity in having monovalent Fab antigen binders but this can be compensated by selection of suitably tight binders. For a number of applications such

as diagnostics and biosensors it may be preferable to have monovalent Fab fragments. For applications requiring Fc effector functions, the technology already exists for extending the heavy chain gene and expressing the glycosylated whole antibody in mammalian cells.

The ideas presented here address the bottle neck in the identification and evaluation of antibodies. It is now possible to construct and screen at least three orders of magnitude more clones with mono-specificity than previously possible. The potential applications of the method should span basic research and applied sciences.

The foregoing is intended as illustrative of the present invention but not limiting. Numerous variations and modifications can be effected without departing from the true spirit and scope of the invention.

## What Is Claimed Is:

1. A method of producing a linear double stranded DNA expression vector containing a 5' terminal promoter controlling the expression of  $V_H$  and  $V_L$  polypeptides from respective  $V_H$ - and  $V_L$ -coding genes, said  $V_H$  and  $V_L$  polypeptides being capable of forming a heterodimeric receptor of predetermined specificity, which method comprises:

(a) isolating a first diverse population of  $V_H$ -coding genes containing a gene coding for said first  $V_H$  polypeptide, said first population being constituted by linear double stranded DNA molecules defining a restriction site recognized by an endonuclease, said restriction site located 5' terminal to said genes but 3' terminal to a promoter controlling the expression of said  $V_H$ -coding genes;

(b) isolating a second diverse population of  $V_L$ -coding genes containing a gene coding for said  $V_L$  polypeptide, said second population being constituted by linear double stranded DNA molecules defining said restriction site, said restriction site located 3' terminal to said  $V_L$ -coding genes, said  $V_L$ -coding genes located 3' terminal to a promoter controlling the expression of said  $V_L$ -coding genes;

(c) cleaving said double stranded DNA molecules while present in said populations with said endonuclease to produce restriction fragments having a cohesive terminus and either one of said  $V_H$ -coding genes or one of said  $V_L$ -coding genes; and

(d) randomly ligating said restriction fragments one to another via said cohesive terminus of each to produce a diverse population of double stranded linear DNA expression vectors having one of said  $V_H$ -coding genes and one of said  $V_L$ -coding genes in tandem on the same DNA strand and under the control of a single promoter;



(e) isolating from said diverse population of double stranded DNA expression vectors a vector capable of expressing said heterodimeric receptor of predetermined specificity.

5           2.    The method of Claim 1 wherein said  $V_H$ -coding genes are  $V_H$ -coding DNA homologs and said isolating comprises:

10           (a)   subjecting a  $V_H$ -coding gene repertoire to a first primer extension reaction utilizing a first polynucleotide synthesis primer capable of initiating said first reaction by hybridizing to a nucleotide sequence conserved with said repertoire, thereby producing a plurality of different  $V_H$ -coding DNA homolog complements, and subjecting said compliments to a second primer extension reaction utilizing a second polynucleotide synthesis primer capable of initiating said second reaction by hybridizing to a nucleotide sequence conserved among said compliments, thereby producing a plurality of different  $V_H$ -coding DNA homologs; and

20           (b)   operatively linking to an expression vector for expression under the control of a promoter each of a plurality of said different  $V_H$ -coding DNA homologs.

25           3.    The method of Claim 1 wherein said  $V_H$ -coding genes are  $V_H$ -coding DNA homologs and said isolating comprises:

30           (a)   subjecting a complement of a  $V_H$ -coding gene repertoire to a primer extension reaction utilizing a polynucleotide synthesis primer capable of inhibiting said primer extension reaction by hybridizing to a nucleotide sequence conserved among said compliments, thereby producing a plurality of different  $V_H$ -coding DNA homologs; and

          (b)   operatively linking to an expression vector for expression under the control of a promoter each of a plurality of said different  $V_H$ -coding DNA homologs.

35           4.    The method of Claim 1 wherein said  $V_L$ -coding

genes are  $V_L$ -coding DNA homologs and said isolating comprises:

(a) subjecting a  $V_L$ -coding gene repertoire to a first primer extension reaction utilizing a first polynucleotide synthesis primer capable of initiating said first reaction by hybridizing to a nucleotide sequence conserved with said repertoire, thereby producing a plurality of different  $V_L$ -coding DNA homolog complements, and subjecting said compliments to a second primer extension reaction utilizing a second polynucleotide synthesis primer capable of initiating said second reaction by hybridizing to a nucleotide sequence conserved among said compliments, thereby producing a plurality of different  $V_L$ -coding DNA homologs; and

(b) operatively linking to an expression vector for expression under the control of a promoter each of a plurality of said different  $V_L$ -coding DNA homologs.

5. The method of Claim 1 wherein said  $V_L$ -coding genes are  $V_L$ -coding DNA homologs and said isolating comprises:

(a) subjecting a complement of a  $V_L$ -coding gene repertoire to a primer extension reaction utilizing a polynucleotide synthesis primer capable of inhibiting said primer extension reaction by hybridizing to a nucleotide sequence conserved among said compliments, thereby producing a plurality of different  $V_L$ -coding DNA homologs; and

(b) operatively linking to an expression vector for expression under the control of a promoter each of a plurality of said different  $V_L$ -coding DNA homologs.

6. The method of Claim 2 wherein said  $V_L$ -coding genes are  $V_L$ -coding DNA homologs and said isolating comprises:

(a) subjecting a  $V_L$ -coding gene repertoire to a third primer extension reaction utilizing a third polynucleotide synthesis primer capable of initiating said

third reaction by hybridizing to a nucleotide sequence conserved with said repertoire, thereby producing a plurality of different  $V_L$ -coding DNA homolog complements, and subjecting said compliments to a fourth primer extension reaction utilizing a fourth polynucleotide synthesis primer capable of initiating said fourth reaction by hybridizing to a nucleotide sequence conserved among said compliments, thereby producing a plurality of different  $V_L$ -coding DNA homologs; and

(b) operatively linking to an expression vector for expression under the control of a promoter each of a plurality of said different  $V_L$ -coding DNA homologs.

7. The method of Claim 2 wherein said  $V_L$ -coding genes are  $V_L$ -coding DNA homologs and said isolating comprises:

(a) subjecting a complement of a  $V_L$ -coding gene repertoire to a primer extension reaction utilizing a third polynucleotide synthesis primer capable of inhibiting said third primer extension reaction by hybridizing to a nucleotide sequence conserved among said compliments, thereby producing a plurality of different  $V_L$ -coding DNA homologs; and

(b) operatively linking to an expression vector for expression under the control of a promoter each of a plurality of said different  $V_L$ -coding DNA homologs.

8. A cloning system comprising first and second linear double stranded DNA vectors admixed in an aqueous medium together with an endonuclease, each of said vectors defining a promoter, a restriction site cleavable by said endonuclease, and a polylinker, said restriction site being located between said promoter and said polylinker on said first vector, said polylinker being located between said promoter and said restriction site on said second vector, said restriction site being cleavable to produce from each of said vectors respective first and second polylinker-

containing restriction fragments having termini complementary to one another.

9. A method of producing a linear double stranded DNA expression vector containing a 5' terminal promoter controlling the expression of  $V_L$  and  $V_H$  polypeptides from respective  $V_L$ - and  $V_H$ -coding genes, said  $V_L$  and  $V_H$  polypeptides being capable of forming a heterodimeric receptor of predetermined specificity, which method comprises:

(a) isolating a first diverse population of  $V_L$ -coding genes containing a gene coding for said  $V_L$  polypeptide, said first population being constituted by linear double stranded DNA molecules defining a restriction site recognized by an endonuclease, said restriction site located 5' terminal to said genes but 3' terminal to a promoter controlling the expression of said genes;

(b) isolating a second diverse population of  $V_H$ -coding genes containing a gene coding for said  $V_H$  polypeptide, said second population being constituted by linear double stranded DNA molecules defining said restriction site, said restriction site located 3' terminal to said  $V_H$ -coding genes, said  $V_H$ -coding genes located 3' terminal to a promoter controlling the expression of said  $V_H$ -coding genes;

(c) cleaving said double stranded DNA molecules while present in said populations with said endonuclease to produce restriction fragments having a cohesive terminus and either one of said  $V_L$ -coding genes or one of said  $V_H$ -coding genes; and

(d) randomly ligating said restriction fragments one to another via said cohesive terminus of each to produce a diverse population of double stranded linear DNA expression vectors having one of said  $V_H$ -coding genes and one of said  $V_L$ -coding genes in tandem on the same DNA strand and under the control of a single promoter;

(e) isolating from said diverse population of double stranded DNA expression vectors a vector capable of expressing said heterodimeric receptor of predetermined specificity.

5           10. The method of Claim 9 wherein said  $V_L$ -coding genes are  $V_H$ -coding DNA homologs and said isolating comprises:

10           (a) subjecting a  $V_H$ -coding gene repertoire to a first primer extension reaction utilizing a first polynucleotide synthesis primer capable of initiating said first reaction by hybridizing to a nucleotide sequence conserved with said repertoire, thereby producing a plurality of different  $V_H$ -coding DNA homolog complements, and subjecting said compliments to a second primer extension reaction utilizing a second polynucleotide synthesis primer capable of initiating said second reaction by hybridizing to a nucleotide sequence conserved among said compliments, thereby producing a plurality of different  $V_H$ -coding DNA homologs; and

15           (b) operatively linking to an expression vector for expression under the control of a promoter each of a plurality of said different  $V_H$ -coding DNA homologs.

20           11. The method of Claim 9 wherein said  $V_H$ -coding genes are  $V_H$ -coding DNA homologs and said isolating comprises:

25           (a) subjecting a complement of a  $V_H$ -coding gene repertoire to a primer extension reaction utilizing a polynucleotide synthesis primer capable of inhibiting said primer extension reaction by hybridizing to a nucleotide sequence conserved among said compliments, thereby producing a plurality of different  $V_H$ -coding DNA homologs; and

30           (b) operatively linking to an expression vector for expression under the control of a promoter each of a plurality of said different  $V_H$ -coding DNA homologs.

35           12. The method of Claim 9 wherein said  $V_L$ -coding

genes are  $V_L$ -coding DNA homologs and said isolating comprises:

(a) subjecting a  $V_L$ -coding gene repertoire to a first primer extension reaction utilizing a first polynucleotide synthesis primer capable of initiating said first reaction by hybridizing to a nucleotide sequence conserved with said repertoire, thereby producing a plurality of different  $V_L$ -coding DNA homolog complements, and subjecting said compliments to a second primer extension reaction utilizing a second polynucleotide synthesis primer capable of initiating said second reaction by hybridizing to a nucleotide sequence conserved among said compliments, thereby producing a plurality of different  $V_L$ -coding DNA homologs; and

(b) operatively linking to an expression vector for expression under the control of a promoter each of a plurality of said different  $V_L$ -coding DNA homologs.

13. The method of Claim 9 wherein said  $V_L$ -coding genes are  $V_L$ -coding DNA homologs and said isolating comprises:

(a) subjecting a complement of a  $V_L$ -coding gene repertoire to a primer extension reaction utilizing a polynucleotide synthesis primer capable of inhibiting said primer extension reaction by hybridizing to a nucleotide sequence conserved among said compliments, thereby producing a plurality of different  $V_L$ -coding DNA homologs; and

(b) operatively linking to an expression vector for expression under the control of a promoter each of a plurality of said different  $V_L$ -coding DNA homologs.

14. The method of Claim 10 wherein said  $V_H$ -coding genes are  $V_L$ -coding DNA homologs and said isolating comprises:

(a) subjecting a  $V_L$ -coding gene repertoire to a third primer extension reaction utilizing a third polynucleotide synthesis primer capable of initiating said

third reaction by hybridizing to a nucleotide sequence conserved with said repertoire, thereby producing a plurality of different  $V_L$ -coding DNA homolog complements, and subjecting said compliments to a fourth primer extension reaction utilizing a fourth polynucleotide synthesis primer capable of initiating said fourth reaction by hybridizing to a nucleotide sequence conserved among said compliments, thereby producing a plurality of different  $V_L$ -coding DNA homologs; and

(b) operatively linking to an expression vector for expression under the control of a promoter each of a plurality of said different  $V_L$ -coding DNA homologs.

15. The method of Claim 10 wherein said  $V_L$ -coding genes are  $V_L$ -coding DNA homologs and said isolating comprises:

(a) subjecting a complement of a  $V_L$ -coding gene repertoire to a primer extension reaction utilizing a third polynucleotide synthesis primer capable of inhibiting said third primer extension reaction by hybridizing to a nucleotide sequence conserved among said compliments, thereby producing a plurality of different  $V_L$ -coding DNA homologs; and

(b) operatively linking to an expression vector for expression under the control of a promoter each of a plurality of said different  $V_L$ -coding DNA homologs.

ABSTRACT

The present invention relates to a method for isolating from the immunological gene repertoire a gene coding for a receptor having the ability to bind a preselected ligand. Receptors produced by the gene isolated by the method, particularly catalytic receptors, are also contemplated.

5

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000  
1001  
1002  
1003  
1004  
1005  
1006  
1007  
1008  
1009  
1010  
1011  
1012  
1013  
1014  
1015  
1016  
1017  
1018  
1019  
1020  
1021  
1022  
1023  
1024  
1025  
1026  
1027  
1028  
1029  
1030  
1031  
1032  
1033  
1034  
1035  
1036  
1037  
1038  
1039  
1040  
1041  
1042  
1043  
1044  
1045  
1046  
1047  
1048  
1049  
1050  
1051  
1052  
1053  
1054  
1055  
1056  
1057  
1058  
1059  
1060  
1061  
1062  
1063  
1064  
1065  
1066  
1067  
1068  
1069  
1070  
1071  
1072  
1073  
1074  
1075  
1076  
1077  
1078  
1079  
1080  
1081  
1082  
1083  
1084  
1085  
1086  
1087  
1088  
1089  
1090  
1091  
1092  
1093  
1094  
1095  
1096  
1097  
1098  
1099  
1100  
1101  
1102  
1103  
1104  
1105  
1106  
1107  
1108  
1109  
1110  
1111  
1112  
1113  
1114  
1115  
1116  
1117  
1118  
1119  
1120  
1121  
1122  
1123  
1124  
1125  
1126  
1127  
1128  
1129  
1130  
1131  
1132  
1133  
1134  
1135  
1136  
1137  
1138  
1139  
1140  
1141  
1142  
1143  
1144  
1145  
1146  
1147  
1148  
1149  
1150  
1151  
1152  
1153  
1154  
1155  
1156  
1157  
1158  
1159  
1160  
1161  
1162  
1163  
1164  
1165  
1166  
1167  
1168  
1169  
1170  
1171  
1172  
1173  
1174  
1175  
1176  
1177  
1178  
1179  
1180  
1181  
1182  
1183  
1184  
1185  
1186  
1187  
1188  
1189  
1190  
1191  
1192  
1193  
1194  
1195  
1196  
1197  
1198  
1199  
1200  
1201  
1202  
1203  
1204  
1205  
1206  
1207  
1208  
1209  
1210  
1211  
1212  
1213  
1214  
1215  
1216  
1217  
1218  
1219  
1220  
1221  
1222  
1223  
1224  
1225  
1226  
1227  
1228  
1229  
1230  
1231  
1232  
1233  
1234  
1235  
1236  
1237  
1238  
1239  
1240  
1241  
1242  
1243  
1244  
1245  
1246  
1247  
1248  
1249  
1250  
1251  
1252  
1253  
1254  
1255  
1256  
1257  
1258  
1259  
1260  
1261  
1262  
1263  
1264  
1265  
1266  
1267  
1268  
1269  
1270  
1271  
1272  
1273  
1274  
1275  
1276  
1277  
1278  
1279  
1280  
1281  
1282  
1283  
1284  
1285  
1286  
1287  
1288  
1289  
1290  
1291  
1292  
1293  
1294  
1295  
1296  
1297  
1298  
1299  
1300  
1301  
1302  
1303  
1304  
1305  
1306  
1307  
1308  
1309  
1310  
1311  
1312  
1313  
1314  
1315  
1316  
1317  
1318  
1319  
1320  
1321  
1322  
1323  
1324  
1325  
1326  
1327  
1328  
1329  
1330  
1331  
1332  
1333  
1334  
1335  
1336  
1337  
1338  
1339  
1340  
1341  
1342  
1343  
1344  
1345  
1346  
1347  
1348  
1349  
1350  
1351  
1352  
1353  
1354  
1355  
1356  
1357  
1358  
1359  
1360  
1361  
1362  
1363  
1364  
1365  
1366  
1367  
1368  
1369  
1370  
1371  
1372  
1373  
1374  
1375  
1376  
1377  
1378  
1379  
1380  
1381  
1382  
1383  
1384  
1385  
1386  
1387  
1388  
1389  
1390  
1391  
1392  
1393  
1394  
1395  
1396  
1397  
1398  
1399  
1400  
1401  
1402  
1403  
1404  
1405  
1406  
1407  
1408  
1409  
1410  
1411  
1412  
1413  
1414  
1415  
1416  
1417  
1418  
1419  
1420  
1421  
1422  
1423  
1424  
1425  
1426  
1427  
1428  
1429  
1430  
1431  
1432  
1433  
1434  
1435  
1436  
1437  
1438  
1439  
1440  
1441  
1442  
1443  
1444  
1445  
1446  
1447  
1448  
1449  
1450  
1451  
1452  
1453  
1454  
1455  
1456  
1457  
1458  
1459  
1460  
1461  
1462  
1463  
1464  
1465  
1466  
1467  
1468  
1469  
1470  
1471  
1472  
1473  
1474  
1475  
1476  
1477  
1478  
1479  
1480  
1481  
1482  
1483  
1484  
1485  
1486  
1487  
1488  
1489  
1490  
1491  
1492  
1493  
1494  
1495  
1496  
1497  
1498  
1499  
1500  
1501  
1502  
1503  
1504  
1505  
1506  
1507  
1508  
1509  
1510  
1511  
1512  
1513  
1514  
1515  
1516  
1517  
1518  
1519  
1520  
1521  
1522  
1523  
1524  
1525  
1526  
1527  
1528  
1529  
1530  
1531  
1532  
1533  
1534  
1535  
1536  
1537  
1538  
1539  
1540  
1541  
1542  
1543  
1544  
1545  
1546  
1547  
1548  
1549  
1550  
1551  
1552  
1553  
1554  
1555  
1556  
1557  
1558  
1559  
1560  
1561  
1562  
1563  
1564  
1565  
1566  
1567  
1568  
1569  
1570  
1571  
1572  
1573  
1574  
1575  
1576  
1577  
1578  
1579  
1580  
1581  
1582  
1583  
1584  
1585  
1586  
1587  
1588  
1589  
1590  
1591  
1592  
1593  
1594  
1595  
1596  
1597  
1598  
1599  
1600  
1601  
1602  
1603  
1604  
1605  
1606  
1607  
1608  
1609  
1610  
1611  
1612  
1613  
1614  
1615  
1616  
1617  
1618  
1619  
1620  
1621  
1622  
1623  
1624  
1625  
1626  
1627  
1628  
1629  
1630  
1631  
1632  
1633  
1634  
1635  
1636  
1637  
1638  
1639  
1640  
1641  
1642  
1643  
1644  
1645  
1646  
1647  
1648  
1649  
1650  
1651  
1652  
1653  
1654  
1655  
1656  
1657  
1658  
1659  
1660  
1661  
1662  
1663  
1664  
1665  
1666  
1667  
1668  
1669  
1670  
1671  
1672  
1673  
1674  
1675  
1676  
1677  
1678  
1679  
1680  
1681  
1682  
1683  
1684  
1685  
1686  
1687  
1688  
1689  
1690  
1691  
1692  
1693  
1694  
1695  
1696  
1697  
1698  
1699  
1700  
1701  
1702  
1703  
1704  
1705  
1706  
1707  
1708  
1709  
1710  
1711  
1712  
1713  
1714  
1715  
1716  
1717  
1718  
1719  
1720  
1721  
1722  
1723  
1724  
1725  
1726  
1727  
1728  
1729  
1730  
1731  
1732  
1733  
1734  
1735  
1736  
1737  
1738  
1739  
1740  
1741  
1742  
1743  
1744  
1745  
1746  
1747  
1748  
1749  
1750  
1751  
1752  
1753  
1754  
1755  
1756  
1757  
1758  
1759  
1760  
1761  
1762  
1763  
1764  
1765  
1766  
1767  
1768  
1769  
1770  
1771  
1772  
1773  
1774  
1775  
1776  
1777  
1778  
1779  
1780  
1781  
1782  
1783  
1784  
1785  
1786  
1787  
1788  
1789  
1790  
1791  
1792  
1793  
1794  
1795  
1796  
1797  
1798  
1799  
1800  
1801  
1802  
1803  
1804  
1805  
1806  
1807  
1808  
1809  
1810  
1811  
1812  
1813  
1814  
1815  
1816  
1817  
1818  
1819  
1820  
1821  
1822  
1823  
1824  
1825  
1826  
1827  
1828  
1829  
1830  
1831  
1832  
1833  
1834  
1835  
1836  
1837  
1838  
1839  
1840  
1841  
1842  
1843  
1844  
1845  
1846  
1847  
1848  
1849  
1850  
1851  
1852  
1853  
1854  
1855  
1856  
1857  
1858  
1859  
1860  
1861  
1862  
1863  
1864  
1865  
1866  
1867  
1868  
1869  
1870  
1871  
1872  
1873  
1874  
1875  
1876  
1877  
1878  
1879  
1880  
1881  
1882  
1883  
1884  
1885  
1886  
1887  
1888  
1889  
1890  
1891  
1892  
1893  
1894  
1895  
1896  
1897  
1898  
1899  
1900  
1901  
1902  
1903  
1904  
1905  
1906  
1907  
1908  
1909  
1910  
1911  
1912  
1913  
1914  
1915  
1916  
1917  
1918  
1919  
1920  
1921  
1922  
1923  
1924  
1925  
1926  
1927  
1928  
1929  
1930  
1931  
1932  
1933  
1934  
1935  
1936  
1937  
1938  
1939  
1940  
1941  
1942  
1943  
1944  
1945  
1946  
1947  
1948  
1949  
1950  
1951  
1952  
1953  
1954  
1955  
1956  
1957  
1958  
1959  
1960  
1961  
1962  
1963  
1964  
1965  
1966  
1967  
1968  
1969  
1970  
1971  
1972  
1973  
1974  
1975  
1976  
1977  
1978  
1979  
1980  
1981  
1982  
1983  
1984  
1985  
1986  
1987  
1988  
1989  
1990  
1991  
1992  
1993  
1994  
1995  
1996  
1997  
1998  
1999  
2000  
2001  
2002  
2003  
2004  
2005  
2006  
2007  
2008  
2009  
2010  
2011  
2012  
2013  
2014  
2015  
2016  
2017  
2018  
2019  
2020  
2021  
2022  
2023  
2024  
2025  
2026  
2027  
2028  
2029  
2030  
2031  
2032  
2033  
2034  
2035  
2036  
2037  
2038  
2039  
2040  
2041  
2042  
2043  
2044  
2045  
2046  
2047  
2048  
2049  
2050  
2051  
2052  
2053  
2054  
2055  
2056  
2057  
2058  
2059  
2060  
2061  
2062  
2063  
2064  
2065  
2066  
2067  
2068  
2069  
2070  
2071  
2072  
2073  
2074  
2075  
2076  
2077  
2078  
2079  
2080  
2081  
2082  
2083  
2084  
2085  
2086  
2087  
2088  
2089  
2090  
2091  
2092  
2093  
2094  
2095  
2096  
2097  
2098  
2099  
2100  
2101  
2102  
2103  
2104  
2105  
2106  
2107  
2108  
2109  
2110  
2111  
2112  
2113  
2114  
2115  
2116  
2117  
2118  
2119  
2120  
2121  
2122  
2123  
2124  
2125  
2126  
2127  
2128  
2129  
2130  
2131  
2132  
2133  
2134  
2135  
2136  
2137  
2138  
2139  
2140  
2141  
2142  
2143  
2144  
2145  
2146  
2147  
2148  
2149  
2150  
2151  
2152  
2153  
2154  
2155  
2156  
2157  
2158  
2159  
2160  
2161  
2162  
2163  
2164  
2165  
2166  
2167  
2168  
2169  
2170  
2171  
2172  
2173  
2174  
2175  
2176  
2177  
2178  
2179  
2180  
2181  
2182  
2183  
2184  
2185  
2186  
2187  
2188  
2189  
2190  
2191  
2192  
2193  
2194  
2195  
2196  
2197  
2198  
2199  
2200  
2201  
2202  
2203  
2204  
2205  
2206  
2207  
2208  
2209  
2210  
2211  
2212  
2213  
2214



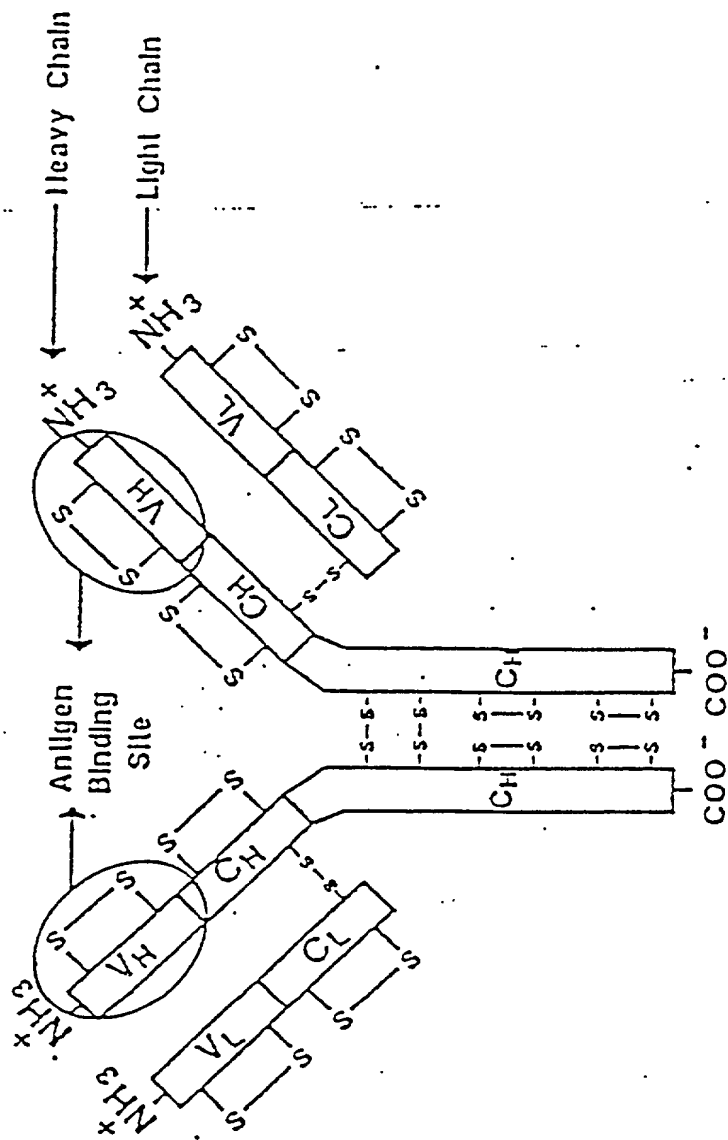
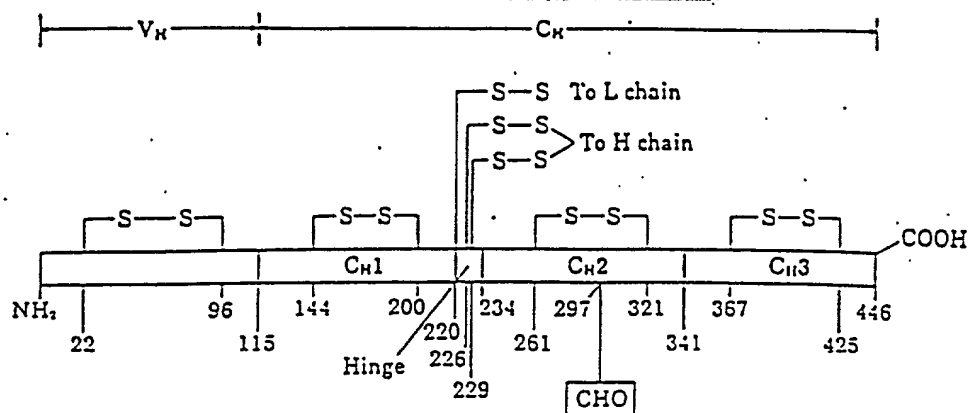


FIGURE 1

A



B

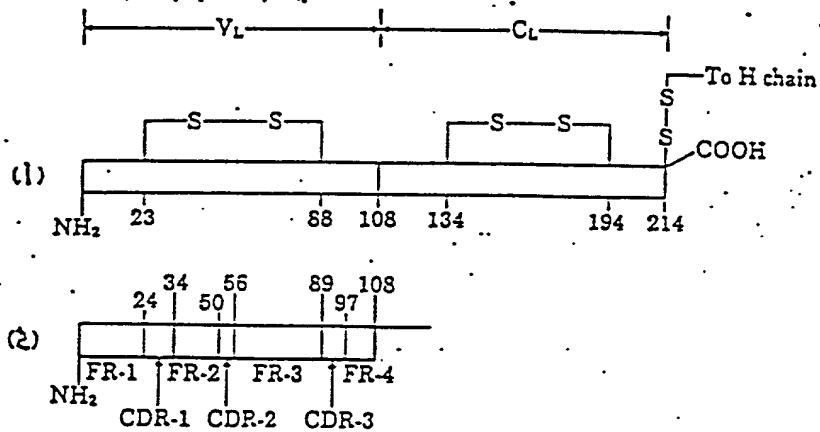


FIGURE 2



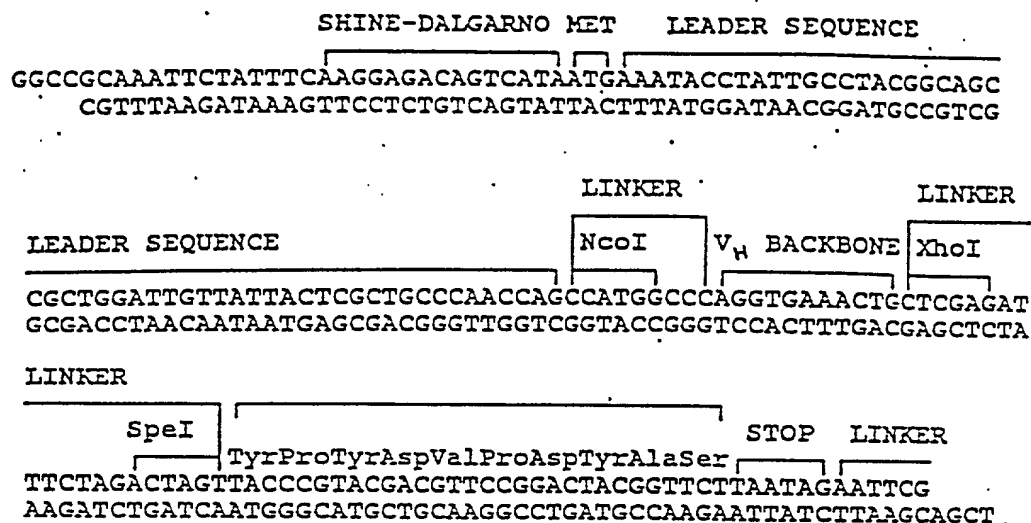


[illegible]

FIGURE 5

A

V<sub>H</sub> EXPRESSION VECTOR:



B

V<sub>L</sub> EXPRESSION VECTOR:

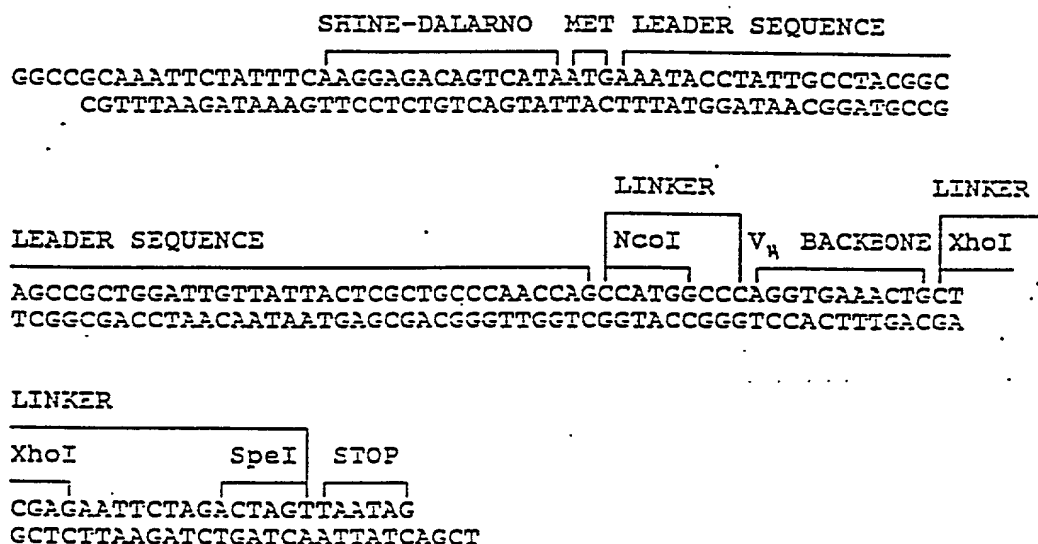


FIGURE 6

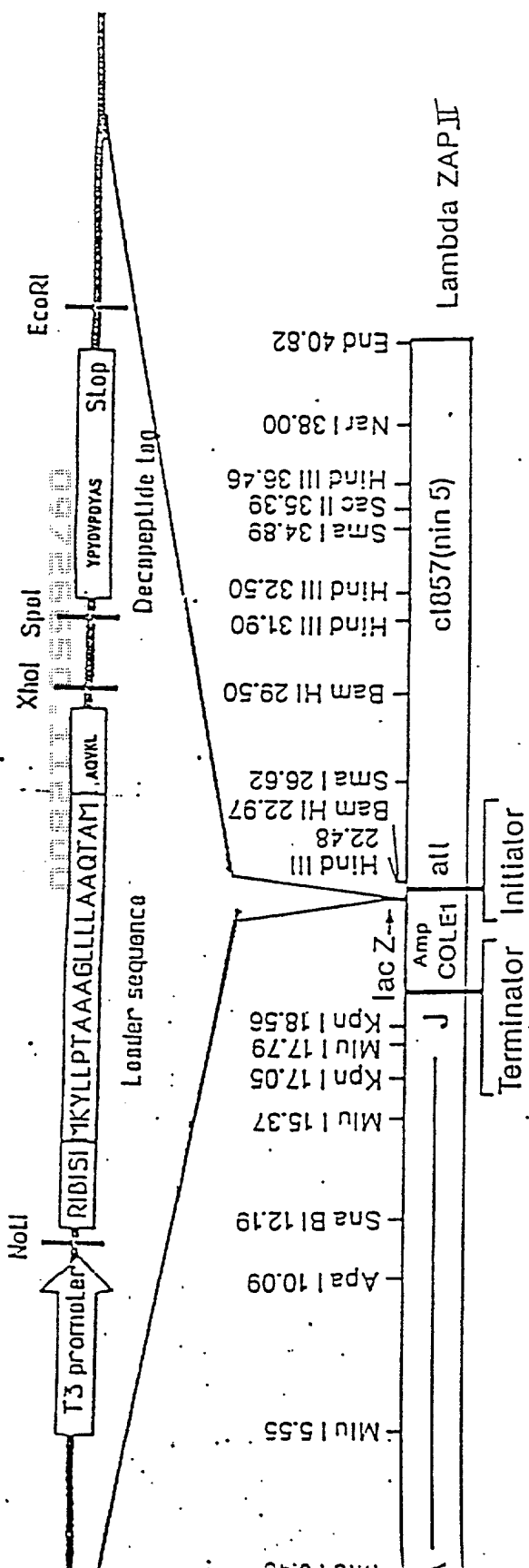


FIGURE 7

Year	Total population		Population aged 15 years and over		Population aged 65 years and over		Population aged 75 years and over		Population aged 85 years and over		Population aged 95 years and over	
	Number	Density	Number	Density	Number	Density	Number	Density	Number	Density	Number	Density
1990	1 000 000	100	700 000	70	100 000	10	50 000	5	25 000	2	10 000	1
2000	1 100 000	110	750 000	75	110 000	11	55 000	5	27 500	2	11 000	1
2010	1 200 000	120	800 000	80	120 000	12	60 000	6	30 000	3	12 000	1
2020	1 300 000	130	850 000	85	130 000	13	65 000	6	32 500	3	13 000	1
2030	1 400 000	140	900 000	90	140 000	14	70 000	7	35 000	3	14 000	1
2040	1 500 000	150	950 000	95	150 000	15	75 000	7	37 500	3	15 000	1
2050	1 600 000	160	1 000 000	100	160 000	16	80 000	8	40 000	4	16 000	1
2060	1 700 000	170	1 050 000	105	170 000	17	85 000	8	42 500	4	17 000	1
2070	1 800 000	180	1 100 000	110	180 000	18	90 000	9	45 000	4	18 000	1
2080	1 900 000	190	1 150 000	115	190 000	19	95 000	9	47 500	4	19 000	1
2090	2 000 000	200	1 200 000	120	200 000	20	100 000	10	50 000	5	20 000	1
2100	2 100 000	210	1 250 000	125	210 000	21	105 000	10	52 500	5	21 000	1

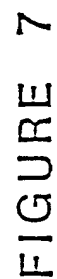


FIGURE 7



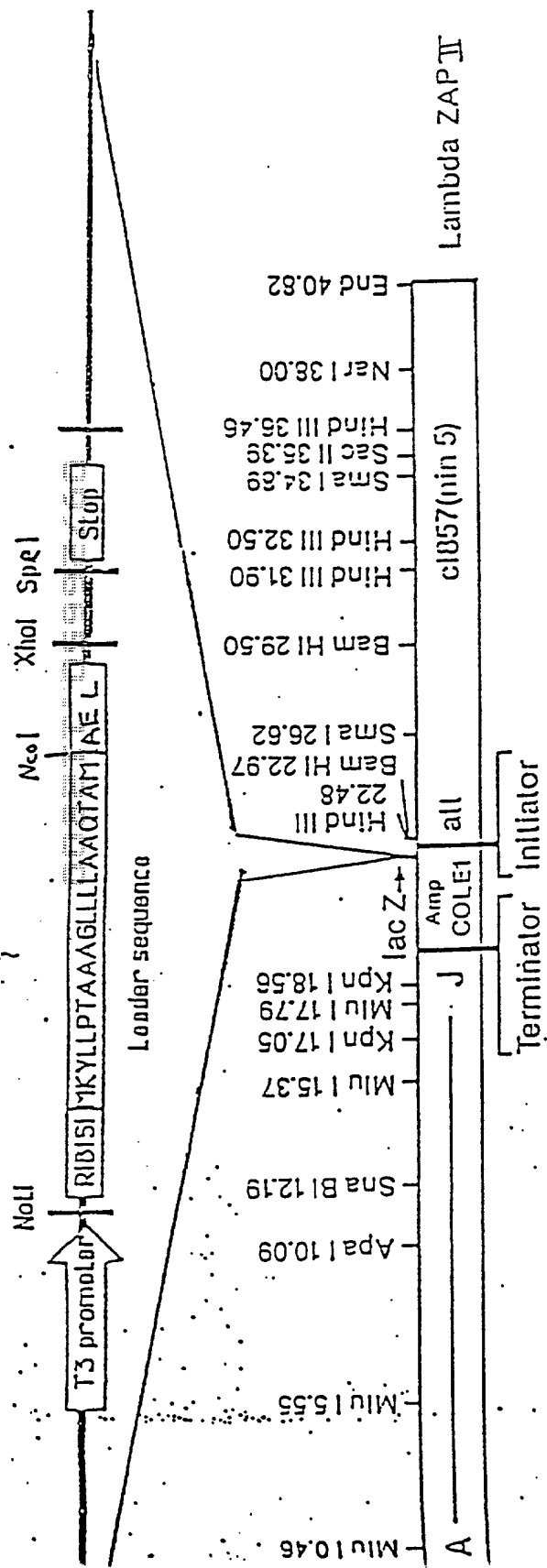


FIGURE 8

[illegible]

FIGURE 8

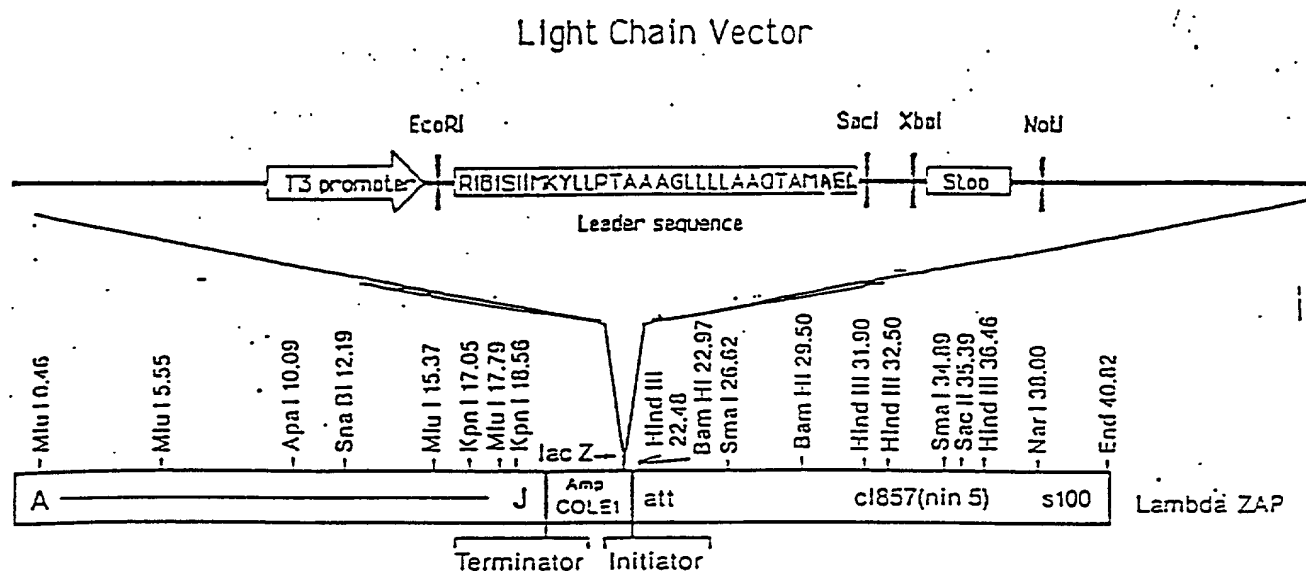


FIGURE 9

EcoR I                      SHINE-DALGARNO    MET                      LEADER SEQUENCE  
 TGAATTCTAACTAGTCGCCAAGGAGACAGTCATAATGAAATACCTATTGCCTACGGCA  
 TCGAACTTAAGATTTGATCAGCGGTTCTCTGTCAGTATTACTTTATGGATAACGGATGCCGT

LEADER SEQUENCE                      Nco I                      Sac I                      Xba I  
 GCCGCTGGATTGTTATTACTCGCTGCCCAACCAGCCATGGCCGAGCTCGTCAGTTCTAGAGT  
 CGGCGACCTAACAATAATGAGCGACGGGTTGGTCGGTACCGGCTCGAGCAGTCAAGATCTCA

Not I  
 TAAGCGGCCG  
 ATTCGCCGGCAGCT

FIGURE 10

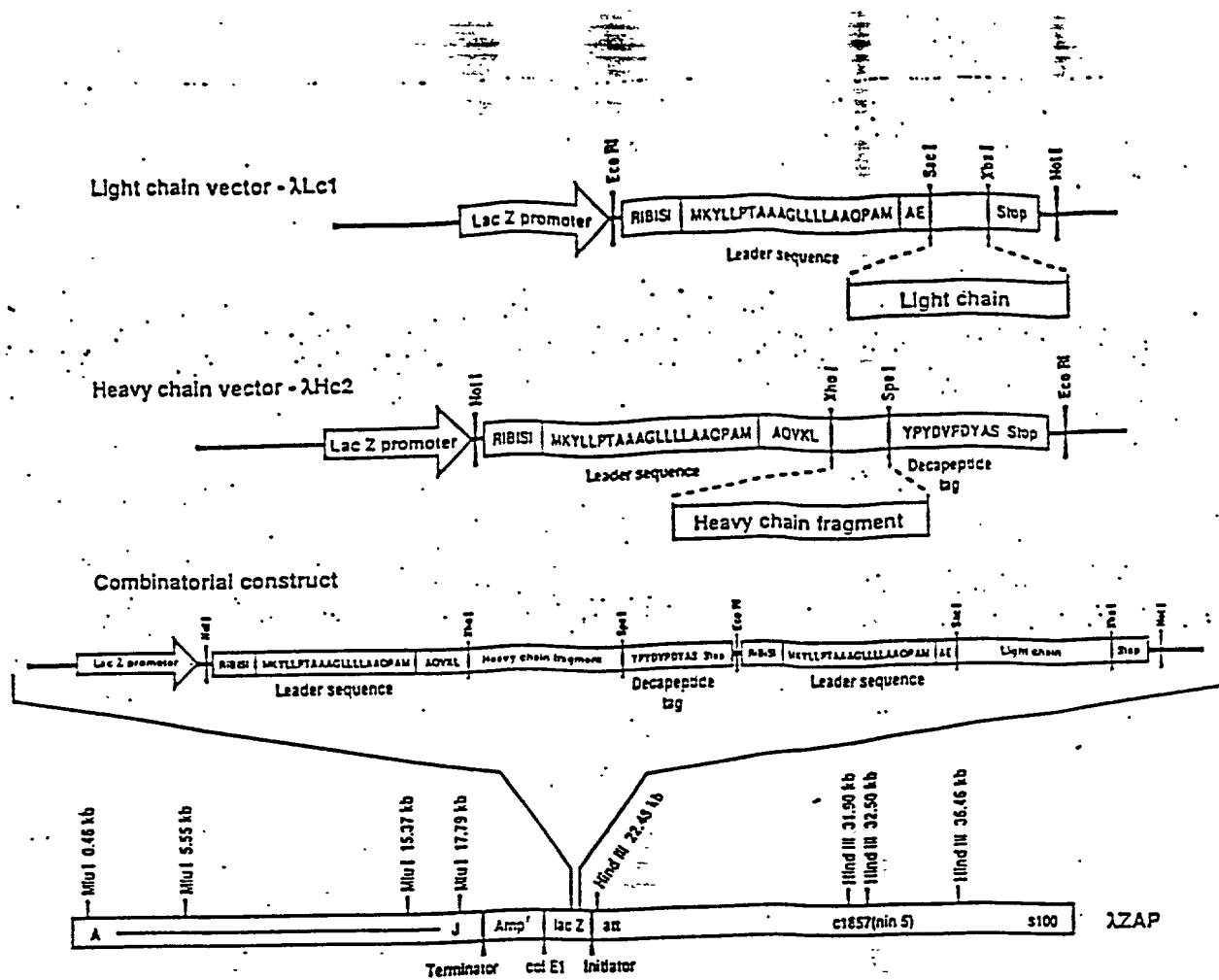


FIGURE 11

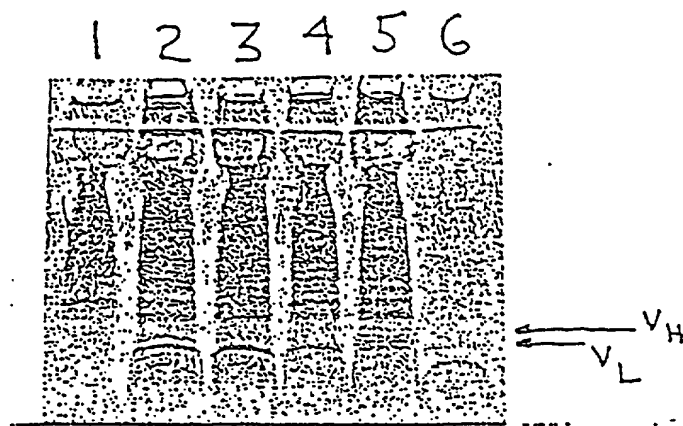


FIGURE 12

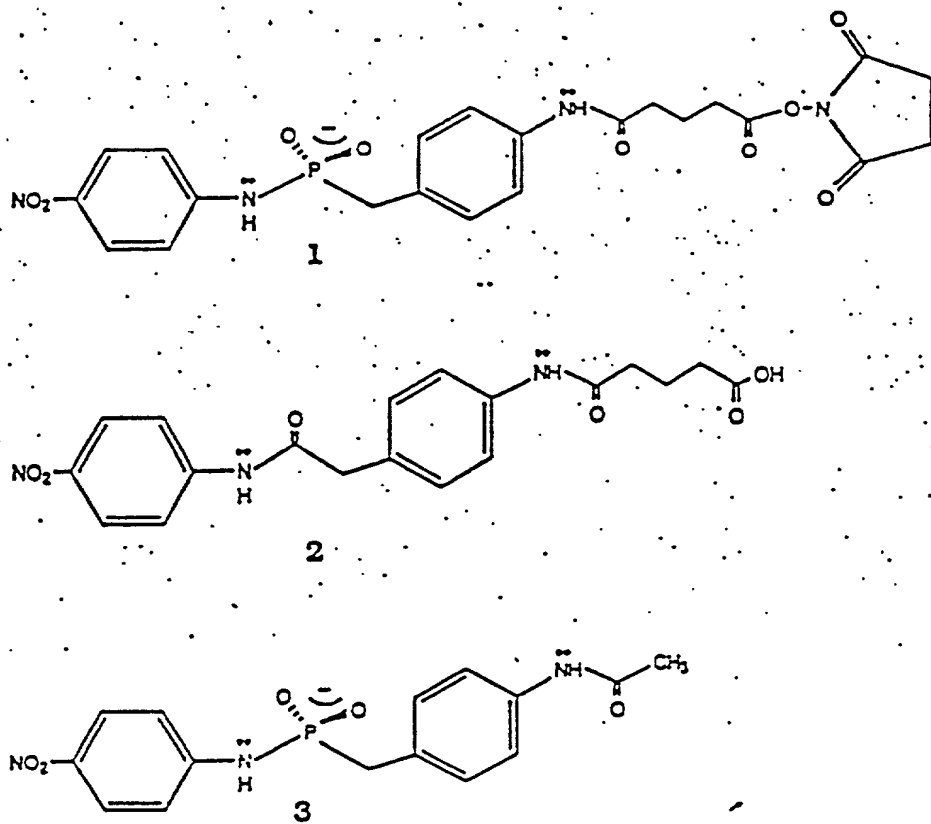


FIGURE 13

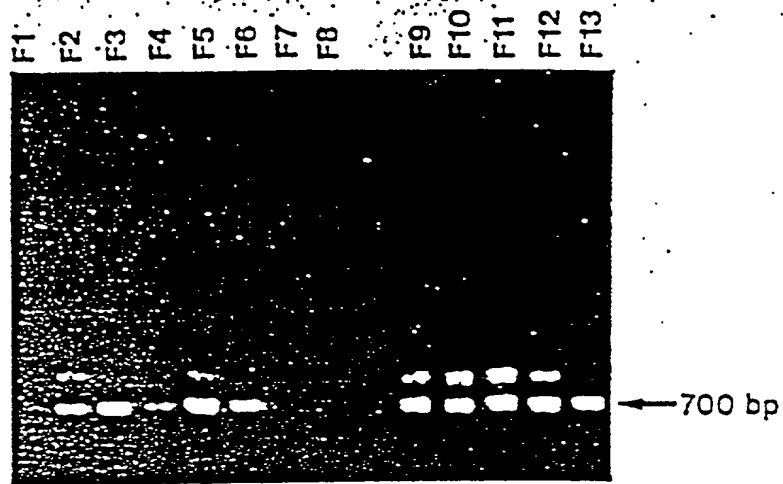


FIGURE 14



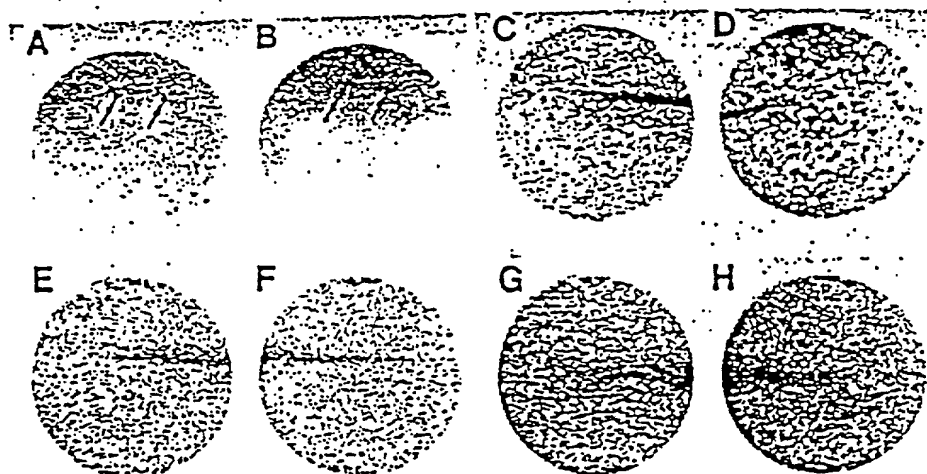


FIGURE 15

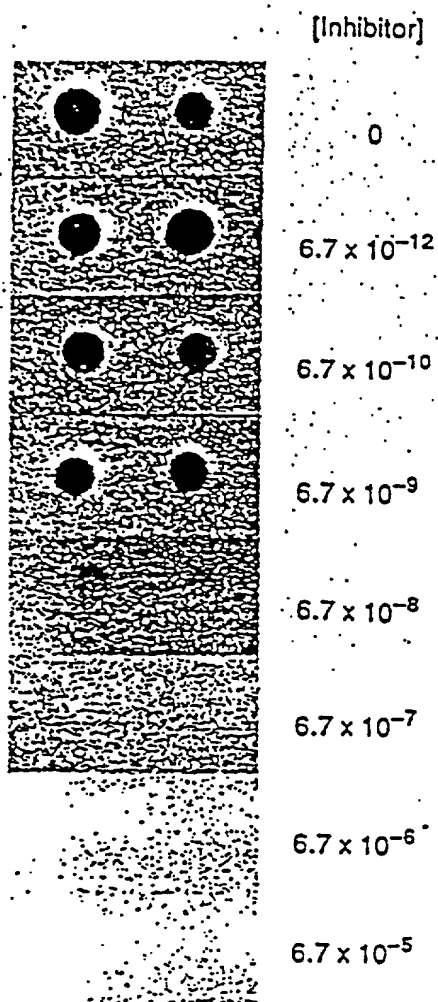


FIGURE 16

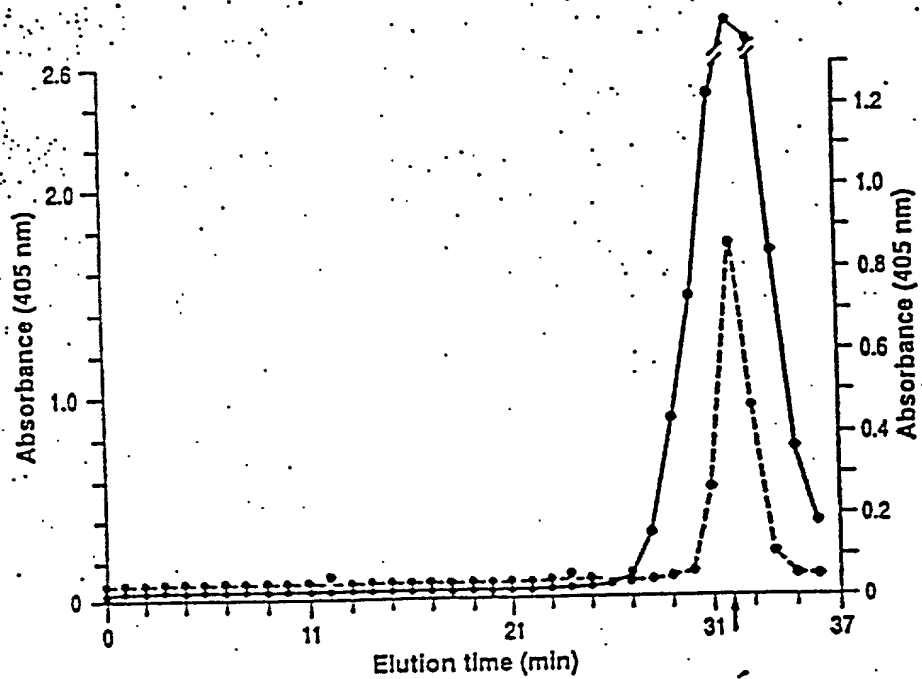


FIGURE 17